

OM nucleic - nucleic search, using sw model  
Run on: September 30, 2003, 18:31:17 ; Search time 14064 Seconds  
(without alignments)  
17362.739 Million cell updates/sec

Title: GS-10-019-284A-1  
Perfect score: 5969  
Sequence: 1 agtccgtgagccaccatt.....tgatatttggccgctgaa 5969

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vl.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5969	100.0	5969	6	BD093238	BD093238 DNA encod
C 2	5056.2	94.8	320550	1	AP005282	AP005282 Corynebac
C 3	5056.2	94.8	349980	6	AX127152	AX127152 Sequence
4	3687	61.8	6911	6	E11760	E11760 Base sequen
5	3687	61.8	6911	6	I26124	I26124 Sequence 4
6	1874.2	31.4	1983	6	AX122988	AX122988 Sequence
7	1874.2	31.4	1983	6	BD165105	BD165105 Novel pol
8	1457	24.4	1656	6	BD094213	BD094213 Genes for
9	1412.6	23.7	1527	6	AX069134	AX069134 Sequence
10	1412.6	23.7	1527	6	AX469820	AX469820 Sequence
11	1321.6	22.1	1342	6	AX065221	AX065221 Sequence
12	1267.8	21.2	1299	6	AX122989	AX122989 Sequence
13	1267.8	21.2	1299	6	BD165106	BD165106 Novel pol
14	1237.4	20.7	1287	6	AX065287	AX065287 Sequence
15	1237.4	20.7	1287	6	AX065289	AX065289 Sequence
16	1110.4	18.6	1152	6	AX122991	AX122991 Sequence
17	1110.4	18.6	1152	6	BD165108	BD165108 Novel pol
18	1049.8	17.6	1109	6	AX069136	AX069136 Sequence
19	762	12.8	882	6	AX065237	AX065237 Sequence
20	671.8	11.3	759	6	AX122990	AX122990 Sequence
21	671.8	11.3	759	6	BD165107	BD165107 Novel pol
22	474.2	7.9	479	6	BD094149	BD094149 Genes for
23	466.6	7.8	498	6	BD094148	BD094148 Genes for
C 24	433	7.3	343050	1	AL935252	AL935252 Lactobaci
C 25	421.6	7.1	21838	1	PDCRAFOPER	PDCRAFOPER
C 26	421.6	7.1	21839	1	PPSURFOF	PPSURFOF
27	397	6.7	5800	1	LL297015	LL297015
28	353.2	5.9	2508	1	STRSCRA	STRSCRA
C 29	352.8	5.9	300330	1	AP005222	AP005222 Corynebac
C 30	351.6	5.9	9979	1	AE015011	AE015011 Streptoco
C 31	325	5.4	2655	1	SXSCRA	X69800 S.xylosus s
C 32	320.2	5.4	10264	1	AE004395	AE004395 Vibrio ch
C 33	320	5.4	2788	1	AF401046	AF401046 Lactobaci
C 34	318.8	5.4	301250	1	AP004830	AP004830 Staphyloc
C 35	318.2	5.3	237850	1	AP003137	AP003137 Staphyloc
C 36	318.2	5.3	341350	1	AP003385	AP003385 Staphyloc
C 37	316.6	5.3	39425	1	AC090968	AC090968 Staphyloc
C 38	303.2	5.1	10085	1	AE006222	AE006222 Pasteural
C 39	301.2	5.0	2955	1	VIBSCRAK	M76768 Vibrio algi
C 40	293.2	4.9	66320	1	AE016833	AE016833 Enterococ
C 41	292.8	4.9	300698	1	AE016750	AE016750 Staphyloc
C 42	281.4	4.7	4158	1	BSTREAPR	254245 B.subtilis
C 43	279.4	4.7	22137	1	D83967	D83967 Bacillus su
C 44	279.4	4.7	37900	1	D86417	D86417 Bacillus su
C 45	279.4	4.7	208430	1	BS080005	Z99108 Bacillus su

ALIGNMENTS

RESULT 1  
BD093238  
LOCUS  
DEFINITION DNA encoding sucrose PTS enzyme II.  
ACCESSION BD093238  
VERSION BD093238.1 GI:22638826  
KEYWORDS WO 0102584-A/1.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 5969)  
AUTHORS Izui,M., Sugimoto,M., Nakamatsu,T. and Kurahashi,O.  
TITLE DNA encoding sucrose PTS enzyme II  
JOURNAL Patent: WO 0102584-A 1 11-JAN-2001;  
AJINOMOTO CO INC,MASAKO IZUI,MASAKAZU SUGIMOTO,TSUYOSHI NAKAMATSU,

OSAMU KURAHASHI	OS	Brevibacterium lactofermentum	PN	WO 0102584-A/1	PD	11-JAN-2001	PF	30-JUN-2000	WO	2000JP004348	PR	02-JUL-1999	JP	99P	189512	PI	MASAKO IZUI,MASAKAZU SUGIMOTO,TSUYOSHI NAKAMATSU,OSAMU PI	KURAHASHI	PC	C12N15/54,C12N9/12	CC		Key	Location/Qualifiers	FT	CDS	Location/Qualifiers	1..5969	/organism="unidentified"	/mol_type="genomic DNA"	/db_xref="taxon:32644"	BASE COUNT	1388 a	1575 c	1578 g	1428 t	ORIGIN	Query Match	100.0%;	Score 5969;	DB 6;	Length 5969;	Best Local Similarity	100.0%;	Pred. No. 0;	Matches 5969;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	Qy	1	AGTCCGTCAGCGCCACCATTCATGCTGCTCAGCGAGCTTGGAGGCTTCTACATCT	60	Db	1	AGTCCGTCAGCGCCACCATTCATGCTGCTCAGCGAGCTTGGAGGCTTCTACATCT	60	Qy	61	ACGCTCCGCTCGCGCTGGAGTGGGTCATTACGGGTGGGATCACGCCGGTGAAGTTGCG	120	Db	61	ACGCTCCGCTCGCGCTGGAGTGGGTCATTACGGGTGGGATCACGCCGGTGAAGTTGCG	120	Qy	121	GAACCCATGGTTCCTTGTGGTGGAGGACGAGTGGGGTGAGAGTTTTCAGATG	180	Db	121	GAACCCATGGTTCCTTGTGGTGGAGGACGAGTGGGGTGAGAGTTTTCAGATG	180	Qy	181	TCTGCAGTTTAAAGTTATGATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG	240	Db	181	TCTGCAGTTTAAAGTTATGATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG	240	Qy	241	CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTTAGACATAGACATATA	300	Db	241	CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTTAGACATAGACATATA	300	Qy	301	CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGTGCATT	360	Db	301	CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGTGCATT	360	Qy	361	ATCAAGAAATGTCAGGTCAAGCTTAAABAAATTTGAGGAGAAATTTTCCCGCTCG	420	Db	361	ATCAAGAAATGTCAGGTCAAGCTTAAABAAATTTGAGGAGAAATTTTCCCGCTCG	420	Qy	421	GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGGAGAAC	480	Db	421	GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGGAGAAC	480	Qy	481	CAGCACCTAAAAACGCGAGGATTCACCCGGAATCCCGAGATGTTCCCGTTTATTG	540	Db	481	CAGCACCTAAAAACGCGAGGATTCACCCGGAATCCCGAGATGTTCCCGTTTATTG	540	Qy	541	ATCTTCATATACCGGTGAACGCTGGCGGTTTCTACGGAACGAGGACCGAGCGGA	600	Db	541	ATCTTCATATACCGGTGGAAACGCTGGCGGTTTCTACGGAACGAGGACCGAGCGGA	600	Qy	601	GGAACCCGCGAGTATACCGGGAACATGGCAGCAGCTGATGTTGCCAAGCATGGTTT	660	Db	601	GGAACCCGCGAGTATACCGGGAACATGGCAGCAGCTGATGTTGCCAAGCATGGTTT	660	Qy	661	CGCGCGGCTGACGACATCGCGCAGGCTGGAAACCTTATCCCTGTTGTGAAGAGG	720	Db	661	CGCGCGGCTGACGACATCGCGCAGGCTGGAAACCTTATCCCTGTTGTGAAGAGG	720	Qy	721	TCCTGCTGTGCGGCAATTCACCTCGAGGGCCCTTTTCATCAACGCATGCGGTTGTGGTCTC	780	Db	721	TCCTGCTGTGCGGCAATTCACCTCGAGGGCCCTTTTCATCAACGCATGCGGTTGTGGTCTC	780	Qy	781	AAAACCCGGATTTCATTTTCCCGGCAACCAACAGATCTTCCCGGGGTGATCCATCGG	840	Db	781	AAAACCCGGATTTCATTTTCCCGGCAACCAACAGATCTTCCCGGGGTGATCCATCGG	840	Qy	841	GAAAAGTTGGATCAATTCGATCAGTAGCGCGGAAACTGACAACTTTCTGAGCTTC	900	Db	841	GAAAAGTTGGATCAATTCGATCAGTAGCGCGGAAACTGACAACTTTCTGAGCTTC	900	Qy	901	TCGATCTCTGCGCAGCCACACATCATTTGCTTCTCGGGCACATGATCAGATTTTG	960	Db	901	TCGATCTCTGCGCAGCCACACATCATTTGCTTCTCGGGCACATGATCAGATTTTG	960	Qy	961	ATACCACATACCAGCGCAATTCGCTTGGCTAAAGAGAAAATGTGACGGTTCAGCGTACGC	1020	Db	961	ATACCACATACCAGCGCAATTCGCTTGGCTAAAGAGAAAATGTGACGGTTCAGCGTACGC	1020	Qy	1021	ATTTGTTCAATCGGATGCTCCGCTGCAATAGGCTCCCGGAGCGTGGCGCTTTGCG	1080	Db	1021	ATTTGTTCAATCGGATGCTCCGCTGCAATAGGCTCCCGGAGCGTGGCGCTTTGCG	1080	Qy	1081	TTGCTGCGGACGCTCCGCGGACGCATATGTTGAGTTGATCGCCGACGCGTGCATTTGG	1140	Db	1081	TTGCTGCGGACGCTCCGCGGACGCATATGTTGAGTTGATCGCCGACGCGTGCATTTGG	1140	Qy	1141	CCGATGGACGCTGATCTAGCTGCTTCCAAACACGCTTTTTCATCGGACGCGATGG	1200	Db	1141	CCGATGGACGCTGATCTAGCTGCTTCCAAACACGCTTTTTCATCGGACGCGATGG	1200	Qy	1201	AAGCCGCCGGAATGCCAGACGGTGTGATACATTTTGGGCGTTTGAACGTCACCGTCACCG	1260	Db	1201	AAGCCGCCGGAATGCCAGACGGTGTGATACATTTTGGGCGTTTGAACGTCACCGTCACCG	1260	Qy	1261	ATGGAGTCCGCGCTGCGGATGGCGGCGCATCGCCGGGGGACACGACACATAGGGA	1320	Db	1261	ATGGAGTCCGCGCTGCGGATGGCGGCGCATCGCCGGGGGACACGACACATAGGGA	1320	Qy	1321	GTCAGTTGTCGACACGCTGCGGAGGGGTATGACGCTTATCGACGGGACCCCTCCACACT	1380	Db	1321	GTCAGTTGTCGACACGCTGCGGAGGGGTATGACGCTTATCGACGGGACCCCTCCACACT	1380	Qy	1381	CAACCGTCGCGCTTAAATTTCTCGGCTTGGCGATCACGAAATCGCTAAATCAAACCTTG	1440	Db	1381	CAACCGTCGCGCTTAAATTTCTCGGCTTGGCGATCACGAAATCGCTAAATCAAACCTTG	1440	Qy	1441	CAAAATTTGTGCTTTTGACTCAACGCGCAGGTGCAAAAGGTGCAATTTAGTTCATCAAG	1500	Db	1441	CAAAATTTGTGCTTTTGACTCAACGCGCAGGTGCAAAAGGTGCAATTTAGTTCATCAAG	1500	Qy	1501	TACTTTAAGTACGAGTAAACTATCTGATTTTAAAGGAGTCCACCATGGAATCACTA	1560	Db	1501	TACTTTAAGTACGAGTAAACTATCTGATTTTAAAGGAGTCCACCATGGAATCACTA	1560	Qy	1561	TCTGCAAAAGACGAGCAAGAAAGTCCGCAAAAGTTCGAGTCTTAATCGCACCTTCGCCA	1620	Db	1561	TCTGCAAAAGACGAGCAAGAAAGTCCGCAAAAGTTCGAGTCTTAATCGCACCTTCGCCA	1620	Qy	1621	ACAAAGGTGGAACCTTGGGCTTGCACAGGATCTTCCACCTGAGTACCTACCAAGAGC	1680	Db	1621	ACAAAGGTGGAACCTTGGGCTTGCACAGGATCTTCCACCTGAGTACCTACCAAGAGC	1680	Qy	1681	TCATTCGCATGTATGAAGCTGGGGAAGTGTTCATTCAAGAACTGCAAGGCAATTTGTTGG	1740	Db	1681	TCATTCGCATGTATGAAGCTGGGGAAGTGTTCATTCAAGAACTGCAAGGCAATTTGTTGG	1740	Qy	1741	ATGAATACGTGGGACTAACCCGTGACGATGAAAACAGCTACTTTAAACCCATTTCGCAAG	1800	Db	1741	ATGAATACGTGGGACTAACCCGTGACGATGAAAACAGCTACTTTAAACCCATTTCGCAAG	1800	Qy	1801	AGTTCACTGACACATCGACATCGTTGATGAAGAGGTCTACAGCCCAAGTGTGTCGAAC	1860	Db	1801	AGTTCACTGACACATCGACATCGTTGATGAAGAGGTCTACAGCCCAAGTGTGTCGAAC	1860
-----------------	----	-------------------------------	----	----------------	----	-------------	----	-------------	----	--------------	----	-------------	----	-----	--------	----	---	-----------	----	--------------------	----	--	-----	---------------------	----	-----	---------------------	---------	--------------------------	-------------------------	------------------------	------------	--------	--------	--------	--------	--------	-------------	---------	-------------	-------	--------------	-----------------------	---------	--------------	---------------	--------------	----	------------	----	--------	----	------	----	----	---	---	----	----	---	---	----	----	----	--	-----	----	----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	--	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	---	-----	----	-----	--	------	----	-----	--	------	----	------	---	------	----	------	---	------	----	------	--	------	----	------	--	------	----	------	---	------	----	------	---	------	----	------	--	------	----	------	--	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	--	------	----	------	--	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	---	------	----	------	--	------	----	------	--	------	----	------	--	------	----	------	--	------	----	------	--	------	----	------	--	------

Db 1801 AGTTCACTGACCACATCGACATCGTTTGATGAAGAGGCTACAGCCAGATGGTGCAAAACC 1860  
Qy 1861 CTGATCCATGCAAGCAGCTGCAGAGTATGAGGCAAGATCCCTGCAGATCCGTTGAAG 1920  
Db 1861 CTGATCCATGCAAGCAGCTGCAGAGTATGAGGCAAGATCCCTGCAGATCCGTTGAAG 1920  
Qy 1921 TTCAAAATCCCTGTGCTCGCGGAAAGCGGACATCGCTTTTCATTTGAACCACTATCTCTCT 1980  
Db 1921 TTCAAAATCCCTGTGCTCGCGGAAAGCGGACATCGCTTTTCATTTGAACCACTATCTCTCT 1980  
Qy 1981 GTGAGGACTGACAAAGGTCAGAGCTGCAGCCTTAACCTGAGGAGCAACGCTCGATT 2040  
Db 1981 GTGAGGACTGACAAAGGTCAGAGCTGCAGCCTTAACCTGAGGAGCAACGCTCGATT 2040  
Qy 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTGTC 2100  
Db 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTGTC 2100  
Qy 2101 CCGCGGCAAAACATCGTTGTTGTCGAAGGAAAGCGGACGCGG 2160  
Db 2101 CCGCGGCAAAACATCGTTGTTGTCGAAGGAAAGCGGACGCGG 2160  
Qy 2161 AACTGTGGAAGGCGGAGTACTCTTCTGCCCAGGTTCCATCTGTAGATGACACACAT 2220  
Db 2161 AACTGTGGAAGGCGGAGTACTCTTCTGCCCAGGTTCCATCTGTAGATGACACACAT 2220  
Qy 2221 GCCACATCATCGTTGATGAAGCAGCAGTATCCAGCTGGAAAGCGCTGATCACTACCG 2280  
Db 2221 GCCACATCATCGTTGATGAAGCAGCAGTATCCAGCTGGAAAGCGCTGATCACTACCG 2280  
Qy 2281 TCTCATGAGCAATTAAGCTCGCTAGAAACAAAAGGAAAGTACTGTGGGGCTATG 2340  
Db 2281 TCTCATGAGCAATTAAGCTCGCTAGAAACAAAAGGAAAGTACTGTGGGGCTATG 2340  
Qy 2341 CACACAGAACTTCCAGTTTGGCCCTGCGTACCATGTGACTCCTCCGAGGCGAGGCTC 2400  
Db 2341 CACACAGAACTTCCAGTTTGGCCCTGCGTACCATGTGACTCCTCCGAGGCGAGGCTC 2400  
Qy 2401 AATGATCCCAACGAATGACGTGATGGAGATACCTCCACCTCTACTACACAGCAGAT 2460  
Db 2401 AATGATCCCAACGAATGACGTGATGGAGATACCTCCACCTCTACTACACAGCAGAT 2460  
Qy 2461 CCAGGTTTCCCTTCGCACCAAGCGCACCGGCTGGCTCACACACACCGCGGTTGACC 2520  
Db 2461 CCAGGTTTCCCTTCGCACCAAGCGCACCGGCTGGCTCACACACACCGCGGTTGACC 2520  
Qy 2521 GGACCGGAGGATGACGTGAGCAGCAGCTCCGAGCTCTTACCCGGATGATCCTAT 2580  
Db 2521 GGACCGGAGGATGACGTGAGCAGCAGCTCCGAGCTCTTACCCGGATGATCCTAT 2580  
Qy 2581 GACCTGGATGATGCTATTCGGTGGAGCGGTATTTACTGACGGCACACTTAACTTTTC 2640  
Db 2581 GACCTGGATGATGCTATTCGGTGGAGCGGTATTTACTGACGGCACACTTAACTTTTC 2640  
Qy 2641 TACACCGGCACTTAAATGACGGAAGCGCGCGCCACCCACCAACCTTGTGAGTC 2700  
Db 2641 TACACCGGCACTTAAATGACGGAAGCGCGCGCCACCCACCAACCTTGTGAGTC 2700  
Qy 2701 GAGGACCAACTGGGCTGATGGGCGGATTCATCGCGTTCGGCTTAAATTCGGCTATC 2760  
Db 2701 GAGGACCAACTGGGCTGATGGGCGGATTCATCGCGTTCGGCTTAAATTCGGCTATC 2760  
Qy 2761 GACGAGCCGCGAGCGGTTTACACCCCATTTACCGGATCCCATGATCAGCCCTGATGT 2820  
Db 2761 GACGAGCCGCGAGCGGTTTACACCCCATTTACCGGATCCCATGATCAGCCCTGATGT 2820  
Qy 2821 GATGGTTGGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTGAGCGGTTCTA 2880  
Db 2821 GATGGTTGGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTGAGCGGTTCTA 2880  
Qy 2881 TACCGCTCGACAGATCTTGAACACTGGGAATTTCTCCGGTGAATACCTTTGACCTCAGT 2940  
Db 2881 TACCGCTCGACAGATCTTGAACACTGGGAATTTCTCCGGTGAATACCTTTGACCTCAGT 2940

Qy 2941 GATGCACAACCTGGTTCTGCTCTCTGATCTCGTCCGATGCTACATGTGGAAATGCCCC 3000  
Db 2941 GATGCACAACCTGGTTCTGCTCTCTGATCTCGTCCGATGCTACATGTGGAAATGCCCC 3000  
Qy 3001 AACCTTTTACGCTTCGGGATGAAGAACTCGGGAAGATCTCGAGCTGCTGATTTCTGT 3060  
Db 3001 AACCTTTTACGCTTCGGGATGAAGAACTCGGGAAGATCTCGAGCTGCTGATTTCTGT 3060  
Qy 3061 CCACAAGGATTGGACCGGAATCCAGATGAGGTACTCTACTACGCAAGCTCTGACCAATGTC 3120  
Db 3061 CCACAAGGATTGGACCGGAATCCAGATGAGGTACTCTACTACGCAAGCTCTGACCAATGTC 3120  
Qy 3121 GGATATGCTGCGACAAGCTTGAAGAAAGCACTTCCGCGCTTTCGAGAGATTCAGCGAG 3180  
Db 3121 GGATATGCTGCGACAAGCTTGAAGAAAGCACTTCCGCGCTTTCGAGAGATTCAGCGAG 3180  
Qy 3181 CTGGATTTCGGGCACTGAATCTTACCCAGCGAGGTGAGTAAACGGTTCTGATGCCCTGG 3240  
Db 3181 CTGGATTTCGGGCACTGAATCTTACCCAGCGAGGTGAGTAAACGGTTCTGATGCCCTGG 3240  
Qy 3241 CTCGTGGCTGGATGGGCTGCCCGCGAGGATGATCACCCAACAGTTGCACAGAAAGA 3300  
Db 3241 CTCGTGGCTGGATGGGCTGCCCGCGAGGATGATCACCCAACAGTTGCACAGAAAGA 3300  
Qy 3301 TGGGTCCACTGCTGACTCTGCCCGCAAGCTTCAATTCGCGAACCCAGCATCTACCA 3360  
Db 3301 TGGGTGCACTGCTGACTCTGCCCGCAAGCTTCAATTCGCGAACCCAGCATCTACCA 3360  
Qy 3361 GAGCTCCTTCTCCAGAGGGGAGTGGGGTAAATCAGATCTGTATTAGTTCTTGAACCT 3420  
Db 3361 GAGCTCCTTCTCCAGAGGGGAGTGGGGTAAATCAGATCTGTATTAGTTCTTGAACCT 3420  
Qy 3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTCTCTG 3480  
Db 3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTCTCTG 3480  
Qy 3481 GATCGTGTGTTGTCGCTGCTGAGTAAACCTGGGAAATAGTATGATCGCGGAC 3540  
Db 3481 GATCGTGTGTTGTCGCTGCTGAGTAAACCTGGGAAATAGTATGATCGCGGAC 3540  
Qy 3541 GATAATACAGCATTGAGATACTGAGTGATGACAGGTTTCATTCGCTTTCCGGG 3600  
Db 3541 GATAATACAGCATTGAGATACTGAGTGATGACAGGTTTCATTCGCTTTCCGGG 3600  
Qy 3601 CTTCAAGGTGACACTATTGAGAGATAAGTCAATAAAGAGGTCCTTTTGGCGAATGT 3660  
Db 3601 CTTCAAGGTGACACTATTGAGAGATAAGTCAATAAAGAGGTCCTTTTGGCGAATGT 3660  
Qy 3661 ACAATACTTCGCAAAATCCCTTGATCGGACACAATAAAGAGTTTATATGTTTAGC 3720  
Db 3661 ACAATACTTCGCAAAATCCCTTGATCGGACACAATAAAGAGTTTATATGTTTAGC 3720  
Qy 3721 TTTTGAACAAACATTCTATCTGATATTTTGGTTTCTCCCGGTTAAGAGAAATTCAT 3780  
Db 3721 TTTTGAACAAACATTCTATCTGATATTTTGGTTTCTCCCGGTTAAGAGAAATTCAT 3780  
Qy 3781 GGACCAATAAGGACCTCGCGCAACGCTCTCGCGACATTCGCGGGAAGACAACATGT 3840  
Db 3781 GGACCAATAAGGACCTCGCGCAACGCTCTCGCGACATTCGCGGGAAGACAACATGT 3840  
Qy 3841 CGCGCGGACACTGTGCAACCGGTTACGCTCTGCTCAAGACACCAAGATGTGGA 3900  
Db 3841 CGCGCGGACACTGTGCAACCGGTTACGCTCTGCTCAAGACACCAAGATGTGGA 3900  
Qy 3901 TCGCCAAAGTCTGGATGATCCAGATCTGAAGGCACTTTTGAACCTGGCGGATGT 3960  
Db 3901 TCGCCAAAGTCTGGATGATCCAGATCTGAAGGCACTTTTGAACCTGGCGGATGT 3960  
Qy 3961 CCAGATCATCTCGGCGGAGGATGTGATCATGTTTCAAGACTCGATGACGCAAC 4020  
Db 3961 CCAGATCATCTCGGCGGAGGATGTGATCATGTTTCAAGACTCGATGACGCAAC 4020

QY 4021 CTCCTCAAGACATCGCTGTCTCCACAGAGCAGCTCAAGATGTTGTGGCTAAACAGCCAA 4080  
|||||  
DB 4021 CTCCTCAAGACATCGCTGTCTCCACAGAGCAGCTCAAGATGTTGTGGCTAAACAGCCAA 4080  
|||||  
QY 4081 CTGGTTACGGCTGCTGTGAAGATATGGCGACATTTTCGTCGCGCTGATTCGAATCTT 4140  
|||||  
DB 4081 CTGGTTACGGCTGCTGTGAAGATATGGCGACATTTTCGTCGCGCTGATTCGAATCTT 4140  
|||||  
QY 4141 GGTGTGGCGGTCTCTCATGGCTATCAACATATGTTGGTTGGCGAGGATCTGTTCGG 4200  
|||||  
DB 4141 GGTGTGGCGGTCTCTCATGGCTATCAACATATGTTGGTTGGCGAGGATCTGTTCGG 4200  
|||||  
QY 4201 TCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGTGATGATCAACCT 4260  
|||||  
DB 4201 TCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGTGATGATCAACCT 4260  
|||||  
QY 4261 GATGGCATCTGCGCGTCTCGCTCTCTGCAAGTGTGTGGTTTTCACCGCAACCAAGCG 4320  
|||||  
DB 4261 GATGGCATCTGCGCGTCTCGCTCTCTGCAAGTGTGTGGTTTTCACCGCAACCAAGCG 4320  
|||||  
QY 4321 TTTCCGTGGCAATGATGTTCCGCGCGCGCATTTGGTATGGGATGGTGTTCGCCAACCCT 4380  
|||||  
DB 4321 TTTCCGTGGCAATGATGTTCCGCGCGCGCATTTGGTATGGGATGGTGTTCGCCAACCCT 4380  
|||||  
QY 4381 GGTAAACGGTACGAGTGGCGCCACCATGACCGCGCGGCAATGCCAATGTGGTCCCT 4440  
|||||  
DB 4381 GGTAAACGGTACGAGTGGCGCCACCATGACCGCGCGGCAATGCCAATGTGGTCCCT 4440  
|||||  
QY 4441 GTTTGGTTGGATGTTGCTCAAGCTGGTTACAGGGCACCGTCTCTGTGCTGGTGGT 4500  
|||||  
DB 4441 GTTTGGTTGGATGTTGCTCAAGCTGGTTACAGGGCACCGTCTCTGTGCTGGTGGT 4500  
|||||  
QY 4501 CTCTTGATTTGGCAACGATCAGAAAGTTCCTGCAAGAGGACTCATGGGACTCAGCA 4560  
|||||  
DB 4501 CTCTTGATTTGGCAACGATCAGAAAGTTCCTGCAAGAGGACTCATGGGACTCAGCA 4560  
|||||  
QY 4561 CTCTTGATCAGCCAGTGTGACTCTGCTGCTACCGGGTCTCTTACGTTGATTCCTAT 4620  
|||||  
DB 4561 CTCTTGATCAGCCAGTGTGACTCTGCTGCTACCGGGTCTCTTACGTTGATTCCTAT 4620  
|||||  
QY 4621 TGTTCAGCAATCGGTGGGTGGGTGACTTGTGCGCACAGGCTCTGACGGGACTCATGA 4680  
|||||  
DB 4621 TGTTCAGCAATCGGTGGGTGGGTGACTTGTGCGCACAGGCTCTGACGGGACTCATGA 4680  
|||||  
QY 4681 TTTCCGTGTCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTACCAATCGTTATCAC 4740  
|||||  
DB 4681 TTTCCGTGTCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTACCAATCGTTATCAC 4740  
|||||  
QY 4741 TGTGTCACAGTCTCTCCGCCAATGAGCTGGAGCTGTTCAACAGGGTGGATCCCT 4800  
|||||  
DB 4741 TGTGTCACAGTCTCTCCGCCAATGAGCTGGAGCTGTTCAACAGGGTGGATCCCT 4800  
|||||  
QY 4801 CATCTTCGCAACCGCATCCATGGCCAAATPCGCGCAGGGTGCAGCATGTTTGGCAGTGT 4860  
|||||  
DB 4801 CATCTTCGCAACCGCATCCATGGCCAAATPCGCGCAGGGTGCAGCATGTTTGGCAGTGT 4860  
|||||  
QY 4861 CTTCCTAGCAGAGTGAAGCTCAAGGCTCAGGCTTGCAGGTGCTCAGGCTCTCCGCTGT 4920  
|||||  
DB 4861 CTTCCTAGCAGAGTGAAGCTCAAGGCTCAGGCTTGCAGGTGCTCAGGCTCTCCGCTGT 4920  
|||||  
QY 4921 TCTTGTGTTACAGAGCTCCGATCTTCGGTGTGAACCTTTCGCTCGCTGGCCGTTCTA 4980  
|||||  
DB 4921 TCTTGTGTTACAGAGCTCCGATCTTCGGTGTGAACCTTTCGCTCGCTGGCCGTTCTA 4980  
|||||  
QY 4981 CATTTGATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTCAGTCTTTGATATCAAGC 5040  
|||||  
DB 4981 CATTTGATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTCAGTCTTTGATATCAAGC 5040  
|||||  
QY 5041 AGTTGCCGTTGGCGCTCAGGTTCTTGGGTGTTGTTCTATTATGATGCTCCAGATATGT 5100  
|||||  
DB 5041 AGTTGCCGTTGGCGCTCAGGTTCTTGGGTGTTGTTCTATTATGATGCTCCAGATATGT 5100  
|||||  
QY 5101 CATGTTCTTGGTTGGCGGTAGTACCTTTTGTATCGCATTCGGCGCAGCGATTCGTTA 5160  
|||||

DB 5101 CATGTTCTTGGTTTGGCGGTAGTTACCTTTGTATCGCATTCGGCGCAGGATTCGTTA 5160  
|||||  
QY 5161 TGGCTTTTACTTGGTTTCGCGCAGCAGCATGATTCAGATGCAACCGCTGCTCCAGT 5220  
|||||  
DB 5161 TGGCTTTTACTTGGTTTCGCGCAGCAGCATGATTCAGATGCAACCGCTGCTCCAGT 5220  
|||||  
QY 5221 GCCTGCAGGAACGACCAAGCCAGAGCAGACCCGCGAGATTTTCAACAGATTCAC 5280  
|||||  
DB 5221 GCCTGCAGGAACGACCAAGCCAGAGCAGACCCGCGAGATTTTCAACAGATTCAC 5280  
|||||  
QY 5281 CATATCCAGCAGCTTGTGACCGGTGAGCTATCGACTGAGCAGCGTTCAGCGATGCCAT 5340  
|||||  
DB 5281 CATATCCAGCAGCTTGTGACCGGTGAGCTATCGACTGAGCAGCGTTCAGCGATGCCAT 5340  
|||||  
QY 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTCGATCGTCCCCACCAAGGGGCGAGCTGT 5400  
|||||  
DB 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTCGATCGTCCCCACCAAGGGGCGAGCTGT 5400  
|||||  
QY 5401 TTCACCAAGTGCAGGAAAGATCGTGGTGGCTTCCCATCTGGTCACGCTTTCGAGTCCG 5460  
|||||  
DB 5401 TTCACCAAGTGCAGGAAAGATCGTGGTGGCTTCCCATCTGGTCACGCTTTCGAGTCCG 5460  
|||||  
QY 5461 CACTAAGGCTGAGGATGGTTCCTCAATGTGGATATCTTGATGCACATTTGGTTTCGACACCGT 5520  
|||||  
DB 5461 CACTAAGGCTGAGGATGGTTCCTCAATGTGGATATCTTGATGCACATTTGGTTTCGACACCGT 5520  
|||||  
QY 5521 AAACCTCAACGCGCAGCCTTTAAACCGCTGAAGAACGAGCGGCGATGAAGTCAAGCAGG 5580  
|||||  
DB 5521 AAACCTCAACGCGCAGCCTTTAAACCGCTGAAGAACGAGCGGCGATGAAGTCAAGCAGG 5580  
|||||  
QY 5581 GGAGTGTGTGTAATTCGATATTTGATGATTCGATTTGATGATTCGATGATGATGATGATGAT 5640  
|||||  
DB 5581 GGAGTGTGTGTAATTCGATATTTGATGATTCGATTTGATGATTCGATGATGATGATGATGAT 5640  
|||||  
QY 5641 GCCGATTTGTTTTCGAAATTAAGAAACCGGACCTGTAACACTTACGCTTTCGCTGGCGA 5700  
|||||  
DB 5641 GCCGATTTGTTTTCGAAATTAAGAAACCGGACCTGTAACACTTACGCTTTCGCTGGCGA 5700  
|||||  
QY 5701 AATTGAAGCGGAGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACACC 5760  
|||||  
DB 5701 AATTGAAGCGGAGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACACC 5760  
|||||  
QY 5761 ATAGCTTGAACCTTGAGTGTTCGACACAGGTTAGACTAGGACGCTGACTCTACGAT 5820  
|||||  
DB 5761 ATAGCTTGAACCTTGAGTGTTCGACACAGGTTAGACTAGGACGCTGACTCTACGAT 5820  
|||||  
QY 5821 CTTTCACACCGGTACCCGTFACGCTTCGAGATTTTAAACCTGTTCAACAGGTCATGCTC 5880  
|||||  
DB 5821 CTTTCACACCGGTACCCGTFACGCTTCGAGATTTTAAACCTGTTCAACAGGTCATGCTC 5880  
|||||  
QY 5881 GGTGATCTGTGTGTGCCACCCCGCATCTTCACCCCGCATTTGGACATGTTGCTTCAGC 5940  
|||||  
DB 5881 GGTGATCTGTGTGTGCCACCCCGCATCTTCACCCCGCATTTGGACATGTTGCTTCAGC 5940  
|||||  
QY 5941 AGTAGCGTTGATATTTTGGCGCGCTGAA 5969  
|||||  
DB 5941 AGTAGCGTTGATATTTTGGCGCGCTGAA 5969  
|||||

RESULT 2  
AP005282/c

LOCUS 320550 bp DNA linear BCT 08-AUG-2002  
DEFINITION Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 9/10  
ACCESSION AP005282 BA000036  
VERSION AP005282.1 GI:21325287  
KEYWORDS  
SOURCE Corynebacterium glutamicum ATCC 13032  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
REFERENCE 1

AUTHORS Nakagawa, S.  
 TITLE Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 320550)  
 AUTHORS Nakagawa, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories: 3-6-6, Asahi-nachi, Machida, Tokyo 194-8533, Japan (E-mail: snakagawa@xanagen.com, Tel: 81-44-829-3031, Fax: 81-44-813-1651)  
 COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.  
 FEATURES  
 Location/Qualifiers  
 1..320550  
 /organism="Corynebacterium glutamicum ATCC 13032"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 13032"  
 /db\_xref="taxon:196627"  
 /note="ATCC 13032"  
 188..814  
 /gene="Cg12516"  
 188..814  
 /gene="Cg12516"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Hypothetical protein"  
 /protein\_id="BAB99909.1"  
 /db\_xref="GI:21325288"  
 /translation="MYRADEFVPLSMVLVTKTRKDHMETLAAQARTLLEKGVAP  
 THAFVESIAKAPILSILLIIVTVNGISGNVQPPALEQVETDVYVKNINERNLK  
 GLVSPLELHTAQTAKRNADSSEKVPDEGLVYLQQLPYANADTIYDR  
 811..2151  
 /gene="Cg12517"  
 811..2151  
 /gene="Cg12517"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Nicotinic acid phosphoribosyltransferase"  
 /protein\_id="BAB99910.1"  
 /db\_xref="GI:21325289"  
 /translation="MNTNPSEFSNRSTALLTKRYELTMQALADGSAERPSTFEVF  
 SRRLPNRRYGVGATVRLKAIKIDFVTEQLADLDFDQRTLEYLRNRYFTGOVVG  
 YREGELVYQSPLLTVRGTEACVILETVILSINNADSAVASAARVTAADGRPIIE  
 MGSRTHEYSAVTASRAYLACFTSTNLEAARYGIPASGTSAAHWTLIHNDDGPV  
 NEAAKFAVSLSDVDTLLVDTYDIQGVATALEVAGPDIGGVIRIDSGDLGVARV  
 RKQDLNNAHTKIYSSDLDEFALAGLRGPDVGVGVTSVTCGAGTAGLVYKIV  
 EVAGHPAKRNKESYGGKKAVRTHRKSGTAIEEIVYFENAEAPDTGKLDLTSLTI  
 PLMRDCEIVGGLP\*LEDSRAYLAKQLVSLPWEGLALSDEPVLHTRFVGFPPAA"  
 2438..2746  
 /gene="Cg12518"  
 2438..2746  
 /gene="Cg12518"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Hypothetical protein"  
 /protein\_id="BAB99911.1"  
 /db\_xref="GI:21325290"  
 /translation="MGLRNLFGPPQKQTQITWKLPAQRNPQAPLSKOSNOTVNRCEA  
 WLSLSTPWLQACLRLNHPQIPQIRSMPLPRKSCILLPLKRNAPVVLDRKWLRL  
 2866..4608  
 /gene="Cg12519"  
 2866..4608  
 /gene="Cg12519"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Rad3-related DNA helicases"  
 /protein\_id="BAB99912.1"  
 /db\_xref="GI:21325291"  
 /translation="MSTATIALQROLVNRDLPLVDALEPLMERRPPTFAIMKGRSNYL  
 CMNKVAROEELNODALIEQEDISWLGHIVRLNEWANETETGDRDLDPGVPLANK  
 QVSVTARECIGASRCPHGDCFAEITARGKAEADVVTNVTHALLAIDALSVDVSVLPEHD

VVVIDAEHLDGRITAVASAEITVNSLNAARRASKLDSKREERVQETAGDLETLIQ  
 TQPGERNMDGSGKTLVALDKDALWALRAQIAGAPEGAANDPERFAERQNLNHLM  
 EHDANRLLEVFAEEDPSKQYDVVNHDDRRGSLNAPLSVALGHLREKLFASNTY  
 VLASATLTIGGNNAWAASWGLPKGSWSDMDAGTDFPAKSLITATRLHPDPGRDGL  
 PEETLDEIYELITAAAGRTGLFSSKRAAEQATKAMRLPLDFVLCQGDNDAAALVKK  
 FDSSENTCYLFTLITQGVYDFGSRSLVIDRIPFPDPDPLLLQARKEADAEGNRG  
 FVEVAAHAALLMAQAGARLLRHVGRGVAVLDHRLTKRYGGFLRFSPRFWETTN  
 PETVRAALKRLVTK"  
 complement(4618..5511)  
 /gene="Cg12520"  
 complement(4618..5511)  
 /gene="Cg12520"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Hypothetical membrane protein"  
 /protein\_id="BAB99913.1"  
 /db\_xref="GI:21325292"  
 /translation="MTSPYGSQYPGDDNNWNSQFNGPSGEQNYGQPYGAPYQPYGVGQ  
 PEDQGFNAYSSPIPEVPPQPSKQEAQWRSFDLGTVEFGMGKFTATQAWVLSALIYF  
 AVLLVLMFAWILPMVSLAATSGSDSAIAATGTGTFEGFMLIVLARISFVYSLNC  
 YRNAARVREGQITIOSFFKMGKLGALGIYILINIVIFIGMILLIPGIIAAVLIIF  
 AVPAFQLRDSAGIDAFSAKSKAVSKNVGVQLLELAIFALSGLSAVIGMLVTTPL  
 TELLYAYASQTASGGPIMORQ"  
 complement(5587..6312)  
 /gene="Cg12521"  
 complement(5587..6312)  
 /gene="Cg12521"  
 /note="PF01981:Domain of unknown function UPF0099"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Uncharacterized ACR"  
 /protein\_id="BAB99914.1"  
 /db\_xref="GI:21325293"  
 /translation="MPKSLDLDAHLKARASDGHSEDPSPSTVOAMQIALHIPKQ  
 QDLPGVTVEDDAKVRAPFLPSLDVSTDPDFAALREYGHILIRKYSRRARNAWDRV  
 NPPRTDVLAAARSVVKLCIDERSVDTPADIKKLOISGTPLDPEINDEYPVIYD  
 ESLKMTLGKAAQVGHASKMLAAHQFFENVEQWAEADFAHVREIPSEEFRLIESPG  
 AVPVROGGTEVAPNTVVAIP"  
 complement(6305..7324)  
 /gene="Cg12522"  
 complement(6305..7324)  
 /gene="Cg12522"  
 /EC\_number="3.1.3.3"  
 /note="PF00702:haloacid dehalogenase-like hydrolase  
 TIGR00099; TIGR00099; conserved hypothetical protein  
 TIGR00338; serfB; phosphoserine phosphatase Serf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Phosphoserine phosphatase"  
 /protein\_id="BAB99915.1"  
 /db\_xref="GI:21325294"  
 /translation="MHGQSVVVELQETVQSSRRSSHVVVLGDPVDALDISRIGOTL  
 ADYDANIDIRGISDYPVGLKLVTPVSPGCGEAMKALALSELNVDIAERS  
 GLLRKRLKVCDFCDSTLITGEVLEMLAHAGKEAYAAVTERAMRGELDFESLRSR  
 VKALAGLDSVIDEAAALIELTGGARTINTLRMGYQTAVSGGQIVLEGLAESE  
 LDYVRANTLEIVDVKLTGNVTIKVIDRAAKAEFLREFAADSGLKMYQTVAVGGDANDI  
 DMLSAAGLGVAFNAPKALKEADTSVNHVFLDEVLHIMGISRDEIDLADQEDGTFRV  
 PLTNA"  
 complement(7811..9565)  
 /gene="Cg12523"  
 complement(7811..9565)  
 /gene="Cg12523"  
 /EC\_number="1.9.3.1"  
 /note="PF00115:Cytochrome C and Quinol oxidase polypeptide  
 I"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Heme/copper-type cytochrome/quinol oxidases,  
 subunit 1"  
 /protein\_id="BAB99916.1"  
 /db\_xref="GI:21325295"  
 /translation="MTAVAPRVGDHVAQPERPTGHARKGSKAWLMTTTHDKQLGIM  
 YIIMSFSEFFLGLGLMALLIRAEFLTQGLFSLNQFNQLFTMGTGTVALLLYGTPVWG

```
FANYVLPAGADVAFPRLNAPGFWITTVGVAMLTGTFPGAADFQWMTYSPUSD
AHSPLGSDMTVGVGATIGSVASAINMLTILCLRAPGTMFMPIFTWNIFFVS
VLALLIFPLLAALGVLDKGLHLYDPANGSLWHLFWFFGHFVYVIALPFF
GIYSEIIPVFSRPMGYVGLIFATISGALSMAWAHMEFTGAVLLPFFSFMFLI
SVPTGVKNWNGTSMKGHTWETPMIWSVPMATPELFGGLTGLMPLSPHLADS
XELLAHFHTLFTGVFASCAGYVFWPMTGRMDERLCKHFNLTFVFGTGLIQ
HWVNGMFRARYADISDGFIIINOISIVFSFLGLSVIPFIWVFKSRWREGELTV
DDPWGNSLEWLEPCPPPHNPFASLPRIISRPAFELHPHMIERKRAEHTGHDD
INAPELGTAPALASDSSR"
9701..10009
/gene="Cg12524"
9701..10009
/gene="Cg12524"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB99917.1"
/db_xref="GI:21325296"
/translators="MPKXVOTIHEVAARESHLWAVTENQCEGVHNLTDISPSPT
PTVNPCHRPNSPRGSIWLRPFQGRFLKXDVPRGPRGAWRSPTICASERPA
"
complement(10086..11090)
/gene="Cg12525"
complement(10086..11090)
/gene="Cg12525"
/note="PF00268:Ribonucleotide reductase, small chain"
/codon_start=1
/transl_table=11
/product="Ribonucleotide reductase beta subunit"
/protein_id="BAB99918.1"
/db_xref="GI:21325297"
/translators="MAADSLVHDYLAKEHVAPVKAINWNSIPDSKDLVWDRLTGN
FWLPKVPVNSDKSMGTNEVEKAATMRVFTGLTLDITQGTVGAISSLPPADSLSHE
EAVLTNIAFMSVHAKSYSNIFMTLASTABIANDAFWSEBENLORAKIILSYEGD
DPLKRTIASVILSEFLYSGFYLPWYSSHSKLTNTADVIRLIRDEAVHGYVIGVY
OKAVNETPRQELKEYFDLLYLDYNETQXSEDLYYDLSWTEYVKRFLRYNANKA
LNKLGEGLFPADETKVPNLSALSPNADENHDFFSGSGSVIVIGKAENTEDDDWDF
"
11555..12043
/gene="Cg12526"
11555..12043
/gene="Cg12526"
/note="PF00210:Ferritin"

Query Match 94.8%; Score 5656.2; DB 1; Length 320550;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

Qy 1 AGTCGGTCGAGCCACCATGATGCTGGTGTCAACGAGCTGCGGAGGCTTCTACATCT 60
Db 152336 AGTCGGTCGAGCCACCATGATGAGTGTGTCAACGAGCTGCGGAGGCTTCTACATCC 152277
Qy 61 ACCTCCCGTCGGCGTGGAGTGGGTCATTACGGGTGGGATCACGCCGGTGAAAGTTGCG 120
Db 152276 ACCTCCCGTCGGCGTGGAGTGGGTCATTACGGGTGGGATCACGCCGGTGAAAGTTGCG 152217
Qy 121 GAACCCATGGTGTCTCTTGGTGTAGGGAAGAGTGGCGGTGAGNAGTTTTCAAGTG 180
Db 152216 GAACCCATGGTGTCTCTTGGTGTAGGGAAGAGTGGCGGTGAGNAGTTTTCAAGTG 152157
Qy 181 TCTCAGTCTTTAAGTTATGCATCATCAGCTTGAAGGCTGAGGCTAATTCAGTAGACCTG 240
Db 152156 TCTCAGTCTTTAAGTTATGCATCATCAGCTTGAAGGCTGAGGCTAATTCAGTAGACCTG 152097
Qy 241 CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAGACATCAT 300
Db 152096 CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAGACATCAT 152037
Qy 301 CGTCCCTATGCTTGGTGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTGCAAT 360
Db 152036 CGTCCCTATGCTTGGTGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTGCAAT 151977
Qy 361 ATCAAGAAATGCAAGGTCAAGCAGTTAAAAAATGAGGGAAGATGTTCCCCCCCTCG 420
```

```
151976 ATCAAGAAATGCAAGGTCAAGCAGTTAAAAAATGAGGGAAGATGTTTACCCCCACG 151917
Qy 421 GGGTATTGATGGCTTTCTCAACTCGAAAAACGCGCATCATCAGGAACCTCTCTGAGAAC 480
Db 151916 GGGTATTGATGGCTTTCTCAACTCGAAAAACGCGCATCATCAGGAACCTCTCTGAGAAC 151857
Qy 481 CAGCACCTAAAAACGACGAGATTCACCCCGCAACTCCCCACGATGTTCCCGGTTTATTG 540
Db 151856 CAGCACCTAAAAACGACGAGATTCACCCCGCAACTCCCCACGATGTTCCCGGTTTATTG 151797
Qy 541 ATCTTCATAATCAGGTGGAAACGCTGGCGGTTCCTACGGGAACGACGACGACGCGA 600
Db 151796 ATCTTCATAATCAGGTGGAAACGCTGGCGGTTCCTACGGGAACGACGACGACGCGA 151737
Qy 601 GGAACACCGCGCAGTATCACCGCGAAATGACACGACGCGTGTATGTCACAGCATGTTT 660
Db 151736 GGAATGCGCGCAGTATCACCGCGAAATGACACGACGCGTGTATGTCACAGCATGTTT 151677
Qy 661 CGCGCCCGGCTGACCACTGCGAGCGAGGTGGAAAACTTATCCCTTGTGTGAAGAGG 720
Db 151676 CGCGCCCGGCTGACCACTGCGAGCGAGGTGGAAAACTTATCCCTTGTGTGAAGAGG 151617
Qy 721 TCCTGCTGTGGCGCATCACCTCGAGGCCCTTTTATCAACGCGATGCGGTTGTGGTGTC 780
Db 151616 GCGTGTGTGGCGCATCACCTCGAGGTCCTTTTATCAACGCGATGCGGTTGTGGTGTC 151557
Qy 781 AAAACCCGATTTTCAATTTTCCCGCAACCAACAGATCTTCCCGGCTGATCCATCGG 840
Db 151556 AAAACCCGATTTTATTTTCCCGCAACCAACAGATCTTCCCGGCTGATCCATCGG 151497
Qy 841 GAAAAGTTGGATCAATTCGATACAGTAGCGCGCGGAAACTGACAATCTTTTACGCTTC 900
Db 151496 GAAAAGTTGGATCAATTCGATACAGTAGCGCGCGGAAACTGACAATCTTTTACGCTTC 151437
Qy 901 TCGATCTCTGCGCAGCGCACACATCATGCTTCTTCTGGGCGACACTGATCAGATTTTG 960
Db 151436 TCGATCTCTGCGCAGCGCACACATCATGCTTCTTCTGGGCGACACTGATCAGATTTTG 151377
Qy 961 ATACCACTACCAAGCGCAATTTGCTTGGCTAAAGAAAAATGTCAGCGTCAACGCTACGC 1020
Db 151376 ATACCACTACCAAGCGCAATTTGCTTGGCTAAAGAAAAATGTCAGCGTCAACGCTACGC 151317
Qy 1021 ATTTGTTCAATCGATGCTCCGCTGTCATCAGGGCTCCCGGAGCGTGGCGGCTTTTCG 1080
Db 151316 ATTTGTTCAATCGATGCTCCGCTGTCATCAGGGATCCCGGAGCGTGGCGGCTTTTCG 151257
Qy 1081 TTGCTGCGCAGCTGCCGGGACGCATATGTTGAGTTGATCGCGACGCGTGCATTTTG 1140
Db 151256 TTGCTGCGCAGCTGCCGGGACGCATATGTTGAGTTGATCGCGACGCGTGCATTTTG 151197
Qy 1141 CCGATGGAACGCTGATCTAGCTCGTTCCAAACAGCGCTTTTTCATCAGGACGCGCATGG 1200
Db 151196 CCGATGGAACGCTGATCTAGCTCGTTCCAAACAGCGCTTTTTCATCAGGACGCGCATGG 151137
Qy 1201 AAGCCGCGGAATGCCAGCGGTGAGTACATTTTGGGGTTTTTGAACGTCAACGTCACCG 1260
Db 151136 AAGCCGCGGAATGCCAGCGGTGAGTACATTTTGGGGTTTTTGAACGTCAACGTCACCG 151077
Qy 1261 ATGAGTGCCTCCGCTCTGGCGATGGCGGCGCATCGCGGGGGACACGACACATAGCGA 1320
Db 151076 ATGAGTGCCTCCGCTCTGGCGATGGCGGCGCATCGCGGGGGTACCAGCACATAGCGA 151017
Qy 1321 GTCAGTTCGTGACACAGCGTGGCAGGGTATGACGCTTATCGACGCGACCTTCCACACT 1380
Db 151016 GTCAGTTCGTGACACAGCGTGGCAGGGTATGACGCTTATCGACGCGACCTTCCACACT 150957
Qy 1381 CAACGCTGCCCTAAAAATCTCGGTCTTGGCGATCAGGAAATCGCTAAATCCAACTG 1440
Db 150956 CAACGCTGCCCTAAAAATCTCGGTCTTGGCGATCAGGAAATCGCTAAATCCAACTG 150897
Qy 1441 GAAATTTGTGCTCTTGTACTCAACGCGCAGGTGCAAAAGTCCATTAGTTCATCAAG 1500
Db 150897 GAAATTTGTGCTCTTGTACTCAACGCGCAGGTGCAAAAGTCCATTAGTTCATCAAG
```

Db	150896	TAAATTTTGTGCTTTTGA	CTCAAAAGCGCCAGTTTACAAAGGTCCTATTTAGACCATCAAG	150837
Qy	1501	TACTTTTAAGTACGAGTAA	AACTATCTCTGAT-TTTAAGGAGTCCCACTATGGAATCACT	1559
Db	150836	TAAATTTAAATACGAGCA	AAACTTTTCTGTATAATAAAGGAGTCCGACCATGGACATCATC	150777
Qy	1560	ATCTGCAAGACGAGCAAG	AGTGGCAAAAGCTGTGCACTCTTAATTCGACACCTTCGGCC	1619
Db	150776	ATCTGCAAGACGAGCAAG	AGTGGCAAAAGCTGTGCACTCTTAATTCGACACCTTCGGCA	150717
Qy	1620	AACAAGGTTGGAACCTT	GGGGCTTGCAACAGGATCTCAACACTAGTACTACCAAGAG	1679
Db	150716	ACTAAGGGCGGAACCTT	GGGGCTTGCAACTGGATGTCACCTTTGAGCACCTACCAAGAG	150657
Qy	1680	CTCATTCGCATGTATGA	AGTGGGAAAGTGTCAATTCAGAACTGCAAGGAGTCTTGTGTTG	1739
Db	150656	CTCATTCGCATGTATGA	AGTGGGAAAGTGTCAATTCAGAACTGCAAGGAGTCTTGTGTTG	150597
Qy	1740	GATGAANTACGTGGGCA	CTAACCCGTGACGATGAATAACAGCTACTTTAAACCATTCGCAAA	1799
Db	150596	GATGAANTACGTGGGAT	TAAAGCGGACGATGAATAACAGCTACTTTAAACCATTCGCTAAA	150537
Qy	1800	GAGTTCATCTACCACT	CGACATCGTTTGATGAAGAGTCTACAGCCAGATGTTGCAAAAC	1859
Db	150536	GAGTTCATCTACCACT	CGACATCGTTTGATGAAGAGTCTACAGCCAGATGTTGCAAAAC	150477
Qy	1860	CCTGATCCATACGAAG	CAGGTGCAAGTATGAGCAAAAGTTCGCTGCAGAACTCGGTTGAA	1919
Db	150476	CCTGATCCATACGAAG	CAGGTGCAAGTATGAGCAAAAGTTCGCTGCAGAACTCGGTTGAT	150417
Qy	1920	GTTCAAAATCCCTTGG	CATCTGGCGGAAACGG-CACATCGCTTTCATTTGAACCATCATCTCT	1978
Db	150416	GTTCAAAATCCCTTGG	CATCTGGCGGAAACGGCACATCGCTTTCATTTGAACCATCATCTCT	150357
Qy	1979	CTGTACGAGTACGAAG	TTCCAGGCTGACCCCTGAACCTTAAAGTGTGGAGGACAGGCTCGA	2038
Db	150356	CTGTACGAGTACGAAG	TTCCAGGCTGACCCCTGAACCTTAAAGTGTGGAGGACAGGCTCGA	150297
Qy	2039	TTCTTCAACACCATCT	GGAAGAGTCCCAACCCACCGCTCACCCAGGTTTGGGACATTTG	2098
Db	150296	TTCTTCAACACCATCT	GGAAGAGTCCCAACCCACCGCTCACCCAGGTTTGGGACATTTG	150237
Qy	2099	TCCGCGCGCAAAACAT	CTGTGGTGGCACTGGTGAAGGAAAGCCGACGCCATCCGC	2158
Db	150236	TCCGCGCGCAAAACAT	CTGTGGTGGCACTGGTGAAGGAAAGCCGACGCCATCCGC	150177
Qy	2159	GGAAGTGGGAAGGCC	CAGTCACTCTTTGCCAGGTTCCATCTGTAGATGCAACAAC	2218
Db	150176	GGAAGTGGGAAGGCC	CAGTCACTCTTTGCCAGGTTCCATCTGTAGATGCAACAAC	150117
Qy	2219	ATGCCACCATCATCGT	TGGATGAAGCAGCAGTATCCAAAGCTGGAATAACCGTGTCACTAC	2278
Db	150057	AATGCCACCATCATCGT	TGGATGAAGCAGCAGTATCCAAAGCTGGAATAACCGTGTCACTAC	150057
Qy	2279	CGTCTCATGGAGCAAT	TAAAGCTCGCTAGAAACAAAGGAAGTACTGTGGGGCTA	2338
Db	150056	CGTCTCATGGAGCAAT	TAAAGCTCGCTAGAAACAAAGGAAGTACTGTGGGGCTA	149997
Qy	2339	TGCACACAGAACTTT	TCCAGTTTGGCCCTCGCTACCATGTGACCTCCCGCAGGGCAGGC	2398
Db	149996	TGCACACAGAACTTT	TCCAGTTTGGCCCTCGCTACCATGTGACCTCCCGCAGGGCAGGC	149937
Qy	2399	TCAATGATCCCAACGG	ATGTACCTGATGAGATACCTCCAGTCTACTTACCAGCAAG	2458
Db	149936	TCAATGATCCCAACGG	ATGTACCTGATGAGATACCTCCAGTCTACTTACCAGCAAG	149877
Qy	2459	ATCCAGGTTTCCCTT	TGCACAAAGGCGACCGGCTGGGCTGCACACCACCGCGCTTGA	2518
Db	149876	ATCCAGGTTTCCCTT	TGCACAAAGGCGACCGGCTGGGCTGCACACCACCGCGCTTGA	149817
Qy	2519	CCGAGCCGACGATTT	GCAGTGGAGCAGCTGCGCGGAGCTTTTACCCGAGTGCATCCT	2578
Db	149816	CCGAGCCGACGATTT	GCAGTGGAGCAGCTGCGCGGAGCTTTTACCCGAGTGCATCCT	149757
Qy	2579	ATGACCTGATGATGCT	ATTTCCGGTGGAGCGGTATTTACTGACGGCACACTTAAACTTT	2638
Db	149756	ATGACCTGATGATGCT	ATTTCCGGTGGAGCGGTATTTACTGACGGCACACTTAAACTTT	149697
Qy	2639	TCTACACCGGCAACCT	AAAAATTGACGGAAGCGCGCGCCACCCAAAACCTTTGCGAAG	2698
Db	149696	TCTACACCGGCAACCT	AAAAATTGACGGAAGCGCGCGCCACCCAAAACCTTTGCGAAG	149637
Qy	2699	TGAGGACCCAACTGG	CGCTGATGCGCGCATTCATCGCGGTTGCGCTAAAAATCCGCTTA	2758
Db	149636	TGAGGACCCAACTGG	CGCTGATGCGCGCATTCATCGCGGTTGCGCTAAAAATCCGCTTA	149577
Qy	2759	TGACGCGACCCGCC	AGCGGTTTTCACACCCATTACCGGATCCCATGATCAACCCCTGATG	2818
Db	149576	TGACGCGACCCGCC	AGCGGTTTTCACACCCATTACCGGATCCCATGATCAACCCCTGATG	149517
Qy	2819	GTGATGTTTGGACAT	GTGTTTGGGCGCCACGGAACCTCAACCGTGCAGCGGTTTC	2878
Db	149516	GTGATGTTTGGACAT	GTGTTTGGGCGCCACGGAACCTCAACCGTGCAGCGGTTTC	149457
Qy	2879	TATACCGCTCGACAG	ATCTTGAAACTGCGGTAATCTCCGGTGAATTCACCTTTGACCTCA	2938
Db	149456	TATACCGCTCGACAG	ATCTTGAAACTGCGGTAATCTCCGGTGAATTCACCTTTGACCTCA	149397
Qy	2939	GTGATGCAACACCT	TGCTTCTGCTCTGATCTCGTTCCCGATGGCTACATGTGGGATGCC	2998
Db	149396	GTGATGCAACACCT	TGCTTCTGCTCTGATCTCGTTCCCGATGGCTACATGTGGGATGCC	149337
Qy	2999	CCAAACCTTTTACGCT	TCGCGATGAAGAACTGGGGAAGATCTCGACGCTGCTGATTTCT	3058
Db	149336	CCAAACCTTTTACGCT	TCGCGATGAAGAACTGGGGAAGATCTCGACGCTGCTGATTTCT	149277
Qy	3059	GTCCACAGGATTTGG	ACCAATCCAGATGAGTACTACCTACGCAAGCTCTGACCACT	3118
Db	149276	GTCCACAGGATTTGG	ACCAATCCAGATGAGTACTACCTACGCAAGCTCTGACCACT	149217
Qy	3119	GGCGATATGCTCGA	CAAGCTTGAAGAACACCTTCCCGCTTCTTGGAGATTCAGCG	3178
Db	149216	GGCGATATGCTCGA	CAAGCTTGAAGAACACCTTCCCGCTTCTTGGAGATTCAGCG	149157
Qy	3179	AGCTGATTTGCGCAT	GAATTTCTAGCGACCCAGGTTGCATTAACCGTTCTGATGCT	3238
Db	149156	AGCTGATTTGCGCAT	GAATTTCTAGCGACCCAGGTTGCATTAACCGTTCTGATGCT	149097
Qy	3239	GGCTCTGGGCTGGAT	GGGCTGCCCGCAGGATGATCACCAACAGTTGCACAGGAAG	3298
Db	149096	GGCTCTGGGCTGGAT	GGGCTGCCCGCAGGATGATCACCAACAGTTGCACAGGAAG	149037
Qy	3299	GATGGTGCACTGCT	GATGTCGCCCGCAGGTTTCATTTGGCAACACGCGATCTACC	3358
Db	149036	GATGGTGCACTGCT	GATGTCGCCCGCAGGTTTCATTTGGCAACACGCGATCTACC	148977
Qy	3359	AAGAGCTCCTTCT	CCAGAGGGGAGTCCGGGTAAATCAGATCTGTATTAGTTCTGAAC	3418
Db	148976	AAGAGCTCCTTCT	CCAGAGGGGAGTCCGGGTAAATCAGATCTGTATTAGTTCTGAAC	148917
Qy	3419	CTCTCGAGTAGACAT	CCGAGGCAATTTTCCCTCGAGTGGGATGGTGTCTGTCTG	3478
Db	148916	CTCTCGAGTAGACAT	CCGAGGCAATTTTCCCTCGAGTGGGATGGTGTCTGTCTG	148857
Qy	3479	TGGATCGTGATGCT	CGGTAGCTGAGTAAACCTGCGCAATTTAGTATCGCGG	3538
Db	148856	TGGATCGTGATGCT	CGGTAGCTGAGTAAACCTGCGCAATTTAGTATCGCGG	148797
Qy	3539	ACGATTAATACAGC	CAATTCAGAGTATGACAGGTTTCATTCGCTTTTCCCGG	3598
Db	148796	ACGATTAATACAGC	CAATTCAGAGTATGACAGGTTTCATTCGCTTTTCCCGG	148738
Qy	3599	GCCTTCAAGGTCAC	ACTATTGAGAGTAAAGGCTTTTGGGCGCAAT	3658
Db	148737	GCCTTCAAGGTCAC	ACTATTGAGAGTAAAGGCTTTTGGGCGCAAT	148678

[illegible]

Db 146517 ATCTTTGACACCGGTACCGGTACCGTTCGAGATTTTAAACCTGTTCAACCGGTGTAAGCC 146458  
Qy 5879 TCAGTGTACCTGTGTGTGTCACCGCCGCAATCTTACCCACATTTGGACATGTTCTGTCA 5938  
Db 146457 TCAGTGTACCTGTGTGTGTCACCGCCGCAATCTTACCCACATTTGGACATGTTCTGTCA 146398  
Qy 5939 GCAGTAGCGTTGATATTTTGCCTGCGT 5967  
Db 146397 GCAGTAGCGTTGATATTTTGCCTGCGT 146369

RESULT 3  
AX127152/c 349980 bp DNA linear PAT 11-MAY-2001  
LOCUS  
DEFINITION Sequence 7068 from Patent EP1108790.  
ACCESSION AX127152 AX114121  
VERSION AX127152.1 GI:14041140  
KEYWORDS  
SOURCE  
ORGANISM Corynebacterium glutamicum  
Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE 1  
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7068 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
Location/Qualifiers  
source 1..349980  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3.309.400) split in 11, seq 7068  
2.700.001 3.049.980"  
BASE COUNT 81250 a 97718 c 90621 g 80391 t  
ORIGIN

Query Match 94.8%; Score 5656.2; DB 6; Length 349980;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

Qy 1 AGTCGTCGACCGCCACCACTTATGTTGGTTCACCGAGCTTGGAGGCTTTTCACATCT 60  
Db 115585 AGTCGTCGACCGCCACCACTTATGATGAGTGGTTCACGAGCTTGGAGGCTTTTCACATCC 115526  
Qy 61 ACCTGCCGTCGCGTGGAGTGGGTGATTCAGCGGTGGGATCACGCCGGTGAAGTTGCG 120  
Db 115525 ACATGCGCTCGCGTGGAGTGGGTGATTCAGCGGTGGGATCACGCCGGTGAAGTTGCG 115466  
Qy 121 GAACCCATGCTGCTTGTGTTGAGGACGAGTCCGGTGGAGAGTTTTCAGAGTG 180  
Db 115465 GAAGCCATGCTGCTTGTGTTGAGGACGAGTCCGGTGGAGAGTTTTCAGAGTG 115406  
Qy 181 TGTGAGTTTAAAGTTATGATCATCATGCTTGAAGGCTGAGGTAAATTCAGTACGCTG 240  
Db 115405 TGTGAGTTTAAAGTTATGATCATCATGCTTGAAGGCTGAGGTAAATTCAGTACGCTG 115346  
Qy 241 CAACAGAGGCTCAAGTCCGAGATATTAACCTAGATTCGCTAGACATAGACATCATATA 300  
Db 115345 CAACAGAGGCTCAAGTCCGAGATATTAACCTAGATTCGCTAGACATAGACATCATATA 115286  
Qy 301 CGTCTATGCTGCTGGAAGGACCAATTAACCTCAGAAAGATGCGAGAAGTGGTGCAATT 360  
Db 115285 CGTCTATGCTGCTGGAAGGACCAATTAACCTCAGAAAGATGCGAGAAGTGGTGCAATT 115226  
Qy 361 ATCAAGAAATCAGGTCAAGCAGTTAAATAAATTTAGGGAGAAATTTTCCCGCCCTCG 420  
Db 115225 ATCAAGAAATCAGGTCAAGCAGTTAAATAAATTTAGGGAGAAATTTTCCCGCCCTCG 115166  
Qy 421 GGGTATGATGGCTTTCACCACTCGAAACGGCATCATCAGCAACTCTCTGGAGAC 480  
Db 115165 GGGTATGATGGCTTTCACCACTCGAAACGGCATCATCAGCAACTCTCTGGAGAC 115106

Qy 481 CAGCACTTAAACACCGCAGGATTCACCCCGACTCCCGCAGATGTTCCCGGTTTATTG 540  
Db 115105 CAGCACTTAAACACCGCAGGATTCACCCCGACTCCCGCAGATGTTCCCGGTTTATTG 115046  
Qy 541 ATCTTCATAATACACCGTGGAAACGGTGGCGGTTTCTTACGAGAACGACGAGCAGCGA 600  
Db 115045 ATCTTCATAATACACCGTGGAAACGGTGGCGGTTTCTTACGAGAACGACGAGCAGCGA 114986  
Qy 601 GGAACACCGCGCAGTATCACCGGACATGCGACCGATGATGTTGCAAGCATGGTTT 660  
Db 114985 GGAATGCGCGCAGTATCACCGGACATGCGACCGATGATGTTGCAAGCATGGTTT 114926  
Qy 661 CGCGCCCGGCTGACGCATTCGCGAGCGCAGGTGGAAACCTTATTCCTTGTGTGAAGAG 720  
Db 114925 CGCGCCCGGCTGACGCATTCGCGAGCGCAGGTGGAAACCTTATTCCTTGTGTGAAGAG 114866  
Qy 721 TCTGCTGTGCGGCTTACCTCGAGGCGCTTTCATCAACGCATGCGGTTGTGTGCTC 780  
Db 114865 GCTGCTGTGCGGCTTACCTCGAGGCTTTCATCAACGCATGCGGTTGTGTGCTC 114806  
Qy 781 AAAACCGGATTTCAATTTTCCCGGCAACCCACAGATCTTCCCGGGTGTATCCATCGG 840  
Db 114805 AAAACCGGATTTTATTTTCCCGGCAACCCACAGATCTTCCCGAGGTGATCCATCGG 114746  
Qy 841 GAAAGTTGGATCAATTCGATGACAGTACGAGCGCCGGAACACTCACAATCTTCTGAGCTC 900  
Db 114745 GAAAGTTGGATCAATTCGATGACAGTACGAGCGCCGGAACACTCACAATCTTCTGAGCTC 114686  
Qy 901 TCGATCTCTGCGCAGCGCACCATCATTCATTCCTTCCTTCGCGCACACTGATGAGATTG 960  
Db 114685 TCGATCTCTGCGCAGCGCACCATCATTCATTCCTTCGCGCACACTGATGAGATTG 114626  
Qy 961 ATACCACTACCGAGCGCAATTCGCTGGCTTAAAGAGAAATGACGGTCAAGGCTACGC 1020  
Db 114625 ATACCACTACCGAGCGCAATTCGCTGGCTTAAAGAGAAATGACGGTCAAGGCTACGC 114566  
Qy 1021 ATTTGTTCAATGCGATCGCTCGCTGCATCATAGGCTCCCGGAGCTGGCGCTTTCG 1080  
Db 114565 ATTTGTTCAATGCGATCGCTCGCTGCATCATAGGATCCCGGAGCTGGCGCTTTCG 114506  
Qy 1081 TTCTCGGACGCTGCGGGGAGCATATGTTGAGTTGATTCGCGAGCGGTGCATTTGG 1140  
Db 114505 TTCTCGGACGCTGCGGGGAGCATATGTTGAGTTGATTCGCGAGCGGTGCATTTGG 114446  
Qy 1141 CCATGGAAGCGTGCATTCAGCTGCTTCCAAACACGCTTTTTCATCAGCGAGCGCATGG 1200  
Db 114445 CCATGGAAGCGTGCATTCAGCTGCTTCCAAACACGCTTTTTCATCAGCGAGCGCATGG 114386  
Qy 1201 AAGCCGCGGAATGCCAGACGGTGAATATTTGGGCGTTTGAACGCTACCGCTCACCG 1260  
Db 114385 AAGCCGCGGAATGCCAGACGGTGAATATTTGGGCGTTTGAACGCTACCGCTCACCG 114326  
Qy 1261 ATGGAGTGGCGGCTGCTCGCGATGGCGGCGCATCCCGGGGACACGACACTAGCGA 1320  
Db 114325 ATGGAGTGGCGGCTGCTCGCGATGGCGGCGCATCCCGGGGATCCCGGGGATAGCGA 114266  
Qy 1321 GTCAATTCGTCACCGCTGCGAGGGGTATGACGCTTATTCGAGCGGACCCCTCCACCT 1380  
Db 114265 GTCAATTCGTCACCGCTGCGAGGGGTATGACGCTTATTCGAGCGGACCCCTCCACCT 114206  
Qy 1381 CAACCGTGGCGGTAAATTTCTCGGCTTGGGATCAGAAATCGCTAAATCCAAACCTG 1440  
Db 114205 CAACCGTGGCGGTAAATTTCTCGGCTTGGGATCAGAAATCGCTAAATCCAAACCTG 114146  
Qy 1441 CAATTTTGTGCTTGTGACTCAACCGCGAGGTGCAAAAGGTCCATTTAGGTCAATCAAG 1500  
Db 114145 TAAATTTGTGCTTGTGACTCAACCGCGAGGTGCAAAAGGTCCATTTAGACCATCAAG 114086  
Qy 1501 TACTTTAAGTACGAGTAAATTCCTGAT-TTTAAAGAGTCCCACTGGAATCACT 1559  
Db 114085 TAAATTTAAGTACGAGTAAATTCCTGAT-TTTAAAGAGTCCCACTGGAATCACT 114026



Db	111866	GCCTTTGAACAACATCATCTGCTGATATATTTTGTCTTCTCCGGTTAAGGAAATTC	111807
Qy	3779	ATGGACCATTAAGACCTCGCGCAACGCATCTCGCGACATTCGCGCGCAAGACAAAT	3838
Db	111806	ATGGACCATTAAGACCTCGCGCAACGCATCTCGCGACATTCGCGCGCAAGACAAAT	111747
Qy	3839	GTGCGCGCGCACACACTGTGCAACCGCTTACGCTCTGCTCAAGACACCAAGAGATG	3898
Db	111746	GTGCGCGCGCACACACTGTGCAACCGCTTACGCTCTGCTCAAGACACCAAGAGATG	111687
Qy	3899	GATCGCCAAAGTCTGGATGATGATCAGATCTGAAAGGCACCTTTGAAACTGCGCGCATG	3958
Db	111686	GATCGCCAAAGTCTGGATGATGATCAGATCTGAAAGGCACCTTTGAAACTGCGCGCATG	111627
Qy	3959	TTCCAGATCATCTGCGCGCAGCGGATGCGATCATGTTTCAAGAACTCGATGACGA	4018
Db	111626	TTCCAGATCATCTGCGCGCAGCGGATGCGATCATGTTTCAAGAACTCGATGACGA	111567
Qy	4019	ACCTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGAGATGTTGGCTAAACAACGCC	4078
Db	111566	ACCTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGAGATGTTGGCTAAACAACGCC	111507
Qy	4079	AACCTGGTTACGCGCTGCTGTAAGGTATTCGCGGACATTTTCGTCGCGTGTATCCAAATC	4138
Db	111506	AACCTGGTTACGCGCTGCTGTAAGGTATTCGCGGACATTTTCGTCGCGTGTATCCAAATC	111447
Qy	4139	TTGGTTGGTGGCGCTCTGCTCATGCGTATCAACAATGTTGGTGGTGGCAGGATCTGTTTC	4198
Db	111446	TTGGTTGGTGGCGCTCTGCTCATGCGTATCAACAATGTTGGTGGTGGCAGGATCTGTTTC	111387
Qy	4199	GGTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	4258
Db	111386	GGTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	111327
Qy	4259	CTGATGTCATCTCGCGCTGCGGCTTCTTCCAGTGTGTTGGTGTTCACCGCAACCAAG	4318
Db	111326	CTGATGTCATCTCGCGCTGCGGCTTCTTCCAGTGTGTTGGTGTTCACCGCAACCAAG	111267
Qy	4319	CGTTTCGGTGGCAATGAGTTCCTGGGCGCGGCAATGGATGCGGATGTTGCTGCTGGT	4378
Db	111266	CGTTTCGGGCGCAATGAGTTCCTGGGCGCGGATGGATGCGGATGTTGCTGCTGGT	111207
Qy	4379	CTGGTTAACGGTACGACGTGGCGCCACCATGACCGCGGCAATGCCAATGGTCTC	4438
Db	111206	TTGGTGAAACGGTACGACGTGGCGCCACCATGACCGCGGCAATGCCAATGGTCTC	111147
Qy	4439	CTGTTTGGTTGGATGTTGCTCAAGTGTGTACAGGCGCACCGTGTCTGTGCTGGT	4498
Db	111146	CTGTTTGGTTGGATGTTGCTCAAGTGTGTACAGGCGCACCGTGTCTGTGCTGGT	111087
Qy	4499	GTCTCTTGGATTCCTGGCAACGATCGGAGACTTCTGCAACAGGCACTCATGSGCACTGCA	4558
Db	111086	GTCTCTTGGATTCCTGGCAACGATCGGAGACTTCTGCAACAGGCACTCATGSGCACTGCA	111027
Qy	4559	GACTTCCTGATACCCAGTGTGTGACFTGCTGCTCACCGGCTTCCTAGTTTCATGCT	4618
Db	111026	GACTTCCTGATACCTCAGTGTGACFTGCTGCTCACCGGCTTCCTAGTTTCATGCT	110967
Qy	4619	ATTGGTCCAGCAATGCGTGGTGGGACTTCTGCGCACACCGTCTGACGGCACTAT	4678
Db	110966	ATTGGTCCAGCAATGCGTGGTGGGACTTCTGCGCACACCGTCTGACGGCACTAT	110907
Qy	4679	GATTTCGGTGGTTCAGTTCGCGGCTGCTCTTTCGGTGTGCTTACTCACCAATCGTTATC	4738
Db	110906	GATTTCGGTGGTTCAGTTCGCGGCTGCTCTTTCGGTGTGCTTACTCACCAATCGTATC	110847
Qy	4739	ACTGGTTCGACCAAGTCTTCCGCCAAATGAGCTGGAGCTGTTCACAGGCTGGATCC	4798
Db	110846	ACTGGTTCGACCAAGTCTTCCGCCAAATGAGCTGGAGCTGTTCACAGGCTGGATCC	110787
Qy	4799	TTCACTTCGCAACCGCATCCATGGCCAAATATCGCGAGGCTGCAGCATGTTGGCAGTG	4858
Db	110786	TTCACTTCGCAACCGCATCTATGGCTAATATCGCCCGAGGTCGGGCAITGTTGGCAGTG	110727
Qy	4859	TTCTTCCCTAGCGAAGAGTGAAAAGCTCAAGGCCCTTCAGGTGCTTCAGGTGCTCCGCT	4918
Db	110726	TTCTTCCCTAGCGAAGAGTGAAAAGCTCAAGGCCCTTCAGGTGCTTCAGGTGCTCCGCT	110667
Qy	4919	GTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGGCTGGCGTTC	4978
Db	110666	GTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGGCTGGCGTTC	110607
Qy	4979	TTACATTTGGTATTCGGTACCGCAGCTATCGGTGGCGCTTTGATTCGACTCTTTAATATCAAG	5038
Db	110606	TTACATTTGGTATTCGGTACCGCAGCTATCGGTGGCGCTTTGATTCGACTCTTTAATATCAAG	110547
Qy	5039	GCAGTTTGGCTTGGCGGCTGCGAGTTTCTTGGGTGTGTTTATTGATGCTCCAGATATG	5098
Db	110546	GCAGTTTGGCTTGGCGGCTGCGAGTTTCTTGGGTGTGTTTATTGATGCTCCAGATATG	110487
Qy	5099	GTCAATGTTTGGTGGCGGCTAGTTACCTTTGTCATCGCATTCGGCGCAGCATTCCT	5158
Db	110486	GTCAATGTTTGGTGGCGGCTAGTTACCTTTGTCATCGCATTCGGCGCAGCATTCCT	110427
Qy	5159	TATGGCCTTACTTGGTTTCGCGCGCAACGCGCAGCATTCAGATGCAACGCTGCTCCA	5218
Db	110426	TATGGCCTTACTTGGTTTCGCGCGCAACGCGCAGCATTCAGATGCAACGCTGCTCCA	110367
Qy	5219	GTSCCTGTCAGGAACACCAAGCCGAAGCAGCAAGCACCAGCAAGATTTTCAAAAGATTC	5278
Db	110366	GTSCCTGTCAGGAACACCAAGCCGAAGCAGCAAGCACCAGCAAGATTTTCAAAAGATTC	110307
Qy	5279	ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCAGTGCAGCAGCGTTCAGCATGCC	5338
Db	110306	ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCAGTGCAGCAGCGTTCAGCATGCC	110247
Qy	5339	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTGGGATCGTCCGCCACCAAGGGGAGTG	5398
Db	110246	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTGGGATCGTCCGCCACCAAGGGGAGTG	110187
Qy	5399	GTTTCCACAGTCAGCGGAAAGATCGTGGTGGCCCTTCCCATCTGGTTCAGCCTTTCGACATC	5458
Db	110186	GTTTCCACAGTCAGCGGAAAGATCGTGGTGGCCCTTCCCATCTGGTTCAGCCTTTCGACATC	110127
Qy	5459	CGCCTTAAGGCTGAGGATGGTTCCCAATGTGGATATCTTGTATGCACATTTGGTTCACACC	5518
Db	110126	CGCCTTAAGGCTGAGGATGGTTCCCAATGTGGATATCTTGTATGCACATTTGGTTCACACC	110067
Qy	5519	GTAACCTCAACGGCAGCGACCTTTAACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	5578
Db	110066	GTAACCTCAACGGCAGCGACCTTTAACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	110007
Qy	5579	GGGAGCTGCTGTGTAATTCGATATTTGATGCCATTAAGGCTGCAAGTTATGAGTAAACC	5638
Db	110006	GGGAGCTGCTGTGTAATTCGATATTTGATGCCATTAAGGCTGCAAGTTATGAGTAAACC	109947
Qy	5639	ACGCGATTTGTTTTCGAATTAAGAAACCGGACCTGTAAACACTTACGGTTTGGGCG	5698
Db	109946	ACGCGATTTGTTTTCGAATTAAGAAACCGGACCTGTAAACACTTACGGTTTGGGCG	109887
Qy	5699	GAAATTTGAAGCGGGAGCCAACTGCTCAACGTGCAAGAAAGAAAGAGCGGTGCCAACA	5758
Db	109886	GAAATTTGAAGCGGGAGCCAACTGCTCAACGTGCAAGAAAGAAAGAGCGGTGCCAACA	109827
Qy	5759	CGATAAGTTGAACCTTTGAGTTTCGCAACAGGTTAGACTAGGGGAGCTGACTCTAGC	5818
Db	109826	CGATAAGTTGAACCTTTGAGTTTCGCAACAGGTTAGACTAGGGGAGCTGACTCTAGC	109767
Qy	5819	ATCTTTGACACCGGTACCGGTACGTTTCGAGATTTTAAACCTTTCAACAGGTCATGCC	5878
Db	109766	ATCTTTGACACCGGTACCGGTACGTTTCGAGATTTTAAACCTTTCAACAGGTCATGCC	109707
Qy	5879	TCGGTGTACCTGTGTGGTCCACCGCGCAATCTTCAACCCACATTTGGACATGTTGTTCA	5938
Db	109706	TCGGTGTACCTGTGTGGTCCACCGCGCAATCTTCAACCCACATTTGGACATGTTGTTCA	109647

QY 5939 GCAGTAGCGTTGATATTTTGGCCCGCTG 5967  
|||||  
Db 109646 GCAAGTACGTTGATATTTTGGCCCGCTG 109618  
|||||

RESULT 4  
E11760  
LOCUS E11760 5911 bp DNA linear PAT 29-SEP-1997  
DEFINITION Base sequence of sucrase gene.  
ACCESSION E11760  
VERSION E11760.1 GI:22025396  
KEYWORDS JP 1996196280-A/1.  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1 (bases 1 to 6911)  
Suginohara, Y., Otonari, K., Nagase, K., Tsuchiya, M., Matsui, Y.,  
Yoshinohara, Y., and Nakamatsu, W.  
PATENT: JP 1996196280-A 1 06-AUG-1996;  
AJINOMOTO CO INC  
OS Brevibacterium lactofermentum  
PN JP 1996196280-A/1  
PD 06-AUG-1996  
PF 30-JAN-1995 JP 1995012361  
PI SUGIMOTO MASAKAZU, OTONARI KIYOKO, NAGASE KAZUO, TSUCHIYA  
MAKOTO, PI MATSUI YUTAKA, YOSHINOHARA YASUHIKO, NAKAMATSU WATARU PC  
C12N15/09, C12N1/21, C12P13/06//C12N9/24, (C12N1/21, C12P13/06); CC  
strandedness: Double;  
CC topology: Linear;  
FH key Location/Qualifiers  
FT source 1..6911  
FT /organism='Brevibacterium lactofermentum' FT  
FT /strain='ATCC13869' FT  
FT /plasmid='pSSM30BS' FT  
FT 342..1505 FT  
FT /product='sucrase' FT  
FT N-acetylglucosamine-6-phosphate deacetylase' FT  
FT 2338..3609 FT  
FT CDS  
FT /product='A protein having high homology to FT  
FT N-acetylglucosamine-6-phosphate deacetylase' FT  
FT 4438..5358 FT  
FT /product='A protein high homology to FT  
FT UDP-N-acetylglucosamine-6-phosphate deacetylase' FT  
FT alanyl-D-glutamyl-meso-6-diaminopimelate FT  
FT synthetase' FT  
FT 5570..6577 FT  
FT /product='A protein having homology to FT  
FT phospho-N-acetylglucosamine-6-phosphate FT  
FT transferase' FT  
FT Location/Qualifiers  
FT 1..6911 FT  
FT /organism='Corynebacterium glutamicum' FT  
FT /mol\_type='genomic DNA' FT  
FT /db\_xref='taxon:1718' FT

BASE COUNT 1575 a 1784 c 1892 g 1660 t  
ORIGIN  
Query Match 61.8%; Score 3687; DB 6; Length 6911;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTCCGTCAGCCACCATGATGTTGGTCCACCGAGCTTCGGGAGGCTTTCTACATCT 60  
|||||  
Db 1 AGTCCGTCAGCCACCATGATGTTGGTCCACCGAGCTTCGGGAGGCTTTCTACATCT 60  
|||||

QY 61 AGCTCCCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCACCGCGGTGAAGTTGCG 120  
|||||  
Db 61 AGCTCCCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCACCGCGGTGAAGTTGCG 120  
|||||

QY 121 GAACCCATGTTGTTCTTCTGTTGAGGACGAGTGGGTTGAGAGTTTTCACATG 180  
|||||

121 GAACCCATGTTGTTCTTCTGTTGAGGACGAGTGGGTTGAGAGTTTTCACATG 180  
181 TCTGCAGTTTTTAAGTTATGTCATCATGCTTGAAGGCTGAGGTAAATTCAGTAGACCTG 240  
181 TCTGCAGTTTTTAAGTTATGTCATCATGCTTGAAGGCTGAGGTAAATTCAGTAGACCTG 240  
181 TCTGCAGTTTTTAAGTTATGTCATCATGCTTGAAGGCTGAGGTAAATTCAGTAGACCTG 240  
241 CAACAGAGGCGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATCATA 300  
241 CAACAGAGGCGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATCATA 300  
301 CGTCTATGCTTGTGTTGAAGGACCAATAACCTCAGAAAGATGGCAGAGTGGTGCATT 360  
301 CGTCTATGCTTGTGTTGAAGGACCAATAACCTCAGAAAGATGGCAGAGTGGTGCATT 360  
361 ATCAAGAAATGCGAGGTCACAGCTTAAATAATTTGAGGGAGAAATTTGTCGCCCTCG 420  
361 ATCAAGAAATGCGAGGTCACAGCTTAAATAATTTGAGGGAGAAATTTGTCGCCCTCG 420  
421 GGGTGATTGATGGCTTTCTCCAACTCGAAACGCGCATCATCAGGAACTCTCTGGAGAAC 480  
421 GGGTGATTGATGGCTTTCTCCAACTCGAAACGCGCATCATCAGGAACTCTCTGGAGAAC 480  
481 CAGCACCTTAAACGCGAGGATTCACCCCGAATTCACCGGATTTCCCGGTTTATTG 540  
481 CAGCACCTTAAACGCGAGGATTCACCCCGAATTCACCGGATTTCCCGGTTTATTG 540  
541 ATCTTCATAATCAGCGTGAAGGCTGGCGGTTTCTACGGGAAACGAGGACCGAGCGA 600  
541 ATCTTCATAATCAGCGTGAAGGCTGGCGGTTTCTACGGGAAACGAGGACCGAGCGA 600  
601 GGAACACCGCGAGTATCACC CGGGAACATGCGACGCGTGTATGTTGCCAAGCATGTTT 660  
601 GGAACACCGCGAGTATCACC CGGGAACATGCGACGCGTGTATGTTGCCAAGCATGTTT 660  
661 CGGCGCGGCTGACGACTGCGAGCGAGGTGGAACCTTTATTCCTTGTGTCAAGAGG 720  
661 CGGCGCGGCTGACGACTGCGAGCGAGGTGGAACCTTTATTCCTTGTGTCAAGAGG 720  
721 TCCTGCTGTGGCGGATTCACCTCGAGGCGCTTTTCATCAACGCGATGCGTGTGGTGC 780  
721 TCCTGCTGTGGCGGATTCACCTCGAGGCGCTTTTCATCAACGCGATGCGTGTGGTGC 780  
781 AAAACCCGATTTTCAATTTTCCCGGCAACCCACAGATCTTTCCTGCGGTGATCCATCGG 840  
781 AAAACCCGATTTTCAATTTTCCCGGCAACCCACAGATCTTTCCTGCGGTGATCCATCGG 840  
841 GAAAAGGTTGGATCAAAATCGATCAGTAGCGCGGAAACTGACAATCTTCTGAGCTTC 900  
841 GAAAAGGTTGGATCAAAATCGATCAGTAGCGCGGAAACTGACAATCTTCTGAGCTTC 900  
901 TCGATCTCTGCGAGCGCACCATCATGTTCTTCTGCGGACACATGATGCGAGATTTTG 960  
901 TCGATCTCTGCGAGCGCACCATCATGTTCTTCTGCGGACACATGATGCGAGATTTTG 960  
961 ATACCCTACCGAGCGCAATTCCTTGGCTTAAAGAGAAATAATGTGACGGTACGCTACGC 1020  
961 ATACCCTACCGAGCGCAATTCCTTGGCTTAAAGAGAAATAATGTGACGGTACGCTACGC 1020  
1021 ATTTGTTCAATGCGATGCTCCGCTGATATAGGGCTCCCGGAGCGGTGGGCGCTTTCG 1080  
1021 ATTTGTTCAATGCGATGCTCCGCTGATATAGGGCTCCCGGAGCGGTGGGCGCTTTCG 1080  
1081 TTGCTGGGACGCTGCGGGGACCATATGTTGAGTTGATGCGGACGGGTGATGCGATTTG 1140  
1081 TTGCTGGGACGCTGCGGGGACCATATGTTGAGTTGATGCGGACGGGTGATGCGATTTG 1140  
1141 CCGATGGAACGGTGCATCTAGCTGTTTCCAAACACGCTTTTTCATCACGGAGCGCATG 1200  
1141 CCGATGGAACGGTGCATCTAGCTGTTTCCAAACACGCTTTTTCATCACGGAGCGCATG 1200  
1201 AAGCCCGGGAATCCAGACGGTGTAGTACATTTTGGCGGTTTGAAGCTCACCGTACCG 1260  
1201 AAGCCCGGGAATCCAGACGGTGTAGTACATTTTGGCGGTTTGAAGCTCACCGTACCG 1260

QY	1261	ATGAGTCGCGCTCTGCGGATGCGCGCCGATCGCGGGGACACGACACTAGCGA	1320	QY	2341	CACACAGAACTTTCAGTTTGGCCCTCGTACCATGTGACTCTCCGACGGAGGCTC	2400
Db	1261		1320	Db	2341		2400
QY	1321	ATGAGTCGCGCTCTGCGGATGCGCGCCGATCGCGGGGACACGACACTAGCGA	1380	QY	2401	AATGATCCCAACGGAATGTACGTCGATGAGATACCCCTCCAGCTAGTAGCAGCAGAT	2460
Db	1321		1380	Db	2401		2460
QY	1381	GTGAGTCGTCGACGCTGCGGAGGGGATGACGCTTATCGACGAGCCCTCCACACCT	1440	QY	2461	CCAGCTTTCCTTCGCGACCAAGCGCCGCTGGGCTCACACCACACCGGTTGACC	2520
Db	1381		1440	Db	2461		2520
QY	1441	CAAATTTGTGGCTTTGACTCAAAAGCGGAGGTGCAAAAGTCCATTTAGGTTCATCAAG	1500	QY	2521	GGACCGCAGCGATTCGACGTGGACGACCTGGCCCGACGCTTCTTACCCGAGATGCATCCAT	2580
Db	1441		1500	Db	2521		2580
QY	1501	TACTTTAAGTACGAGTAAACTATCTCTGATTTTAAAGGAGTCCACCATGGAATCACTA	1560	QY	2581	GACCTTGGATGGATGCTATTCGCTGGAGCCGCTTACTGAGGCGACACTTAAACTTTTC	2640
Db	1501		1560	Db	2581		2640
QY	1561	TCTGCAAGACGAGCAAGAGTCGGCAAGCAGTTGACGTCCCTAAATCGCACCCCTTCGCCA	1620	QY	2641	TACACCGGCAACCTTAAATAATTGACGAAAGCGCCGCGCACCCCAAAACCTTGTGAAATC	2700
Db	1561		1620	Db	2641		2700
QY	1621	ACGAGGTGGAACCTTGGGCTTGGCAACAGAGTCTCACACCTGAGTACCTACCAAGAGC	1680	QY	2701	GAGGACCCAACTTGGGCTGATGCGGCGCATTCATCCCGCTTCCCTTAAATAATCCGCTTATC	2760
Db	1621		1680	Db	2701		2760
QY	1681	TCATTCCGATGTGAGCTGGGAGTGTCAATCAAGAGTCTACAGGAGTCTTCTGTGG	1740	QY	2761	GACGACCCCGCAGCGGTTTACACCCCATTACCGGATCCCATGATCAGCCCTCATGGT	2820
Db	1681		1740	Db	2761		2820
QY	1741	ATGAATACGTGGGACTTAACCCGTGACGATGAAACAGCTACTTTAAACCATTCGCAAG	1800	QY	2821	GATGTTGGAACATGGTTCTTGGGGCCCAACCGGAAACCTCACCGGTGACGCGTTCTA	2880
Db	1741		1800	Db	2821		2880
QY	1801	AGTCACTGACCATCGACATCGTTGATGAAGAGTCTACAGCCAGATGTCGAAACC	1860	QY	2881	TACCGCTCGACAGATCTTGAACACTGGGAATCTCCGGTGAATCACCTTTGACCTCAGT	2940
Db	1801		1860	Db	2881		2940
QY	1861	CTGATCCATACGAAGACGTGACAGATGATGAGCAAGATCGGTGAGAAATCCGTTGAAG	1920	QY	2941	GATGCACAACCTTGGCTCTGCTGATCTGCTTCCCGATGGCTACATGCGGAATGCCCC	3000
Db	1861		1920	Db	2941		3000
QY	1921	TTCAATTCCTTGGATCGCGGAAACGGACATCGCTTTCATTTGAACCATCTCTCTCT	1980	QY	3001	AACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATTCTCTGT	3060
Db	1921		1980	Db	3001		3060
QY	1981	GTGAGGACTGACAAAGTCCAGCGGTGACCCCTTAAACTGTGGAGGACAACTCGATT	2040	QY	3061	CCACAAGGATTGGACCGAATCCACGATGAGGTTACTCTACTAGCGAAGCTTACCGAGT	3120
Db	1981		2040	Db	3061		3120
QY	2041	CTTCAACACCATCGAAGAGTCCCAACCCGCGTCACCCAGGTTTGGGCACTTTGTC	2100	QY	3121	GGATATGTCGTCGACAAAGCTTGAAGGAAAGCTTCCGCTTCCGAGGATTCAGCGAG	3180
Db	2041		2100	Db	3121		3180
QY	2101	CCGCGCAAAACATCGTTGTGTGGCAACTGTGAAAGGAAAGCGACGCCATCCGCGG	2160	QY	3181	CTGGATTTCGGCCATGAATTTACGACACCGCAGGTTGAGTAAACGGTTCTGATGCCCTGG	3240
Db	2101		2160	Db	3181		3240
QY	2161	AACCTGAGAGGCCAGTGTCTTCTGCCCAGGTTCCATCTGTAGATGACAAACAT	2220	QY	3241	CTCGTGGCTGGATGGGCTGCCCGCGAGGATGATACCCAACAGTTGCACAGGAAGA	3300
Db	2161		2220	Db	3241		3300
QY	2221	GCCACCATCATCGTTGATGAAGCAGCAGTATCCAAGCTGGAACACGCTGATCAGTACCG	2280	QY	3301	TGGGTGCACTGCTGACTGTGCCCGCGCAAGCTTCAATTTGCGCAACACGCGATCTACCAA	3360
Db	2221		2280	Db	3301		3360
QY	2281	TCTCATGAGCAATTAAGCTGCGCTAGAAACAAAGGAAAGTACTGTGTGGGGTATG	2340	QY	3361	GAGCTCTTCTCCAGAGGGGAGTGGGGGTAAATCAGATCTGTATTAGGTTCTGAACT	3420
Db	2281		2340	Db	3361		3420
QY	2341	GTCCGAGTAGACATCCGAGGCAATATTTCCTCCGAGTGGGATGGTCCGTTTGTCTGTG	3480	QY	3421	GTCCGAGTAGACATCCGAGGCAATATTTCCTCCGAGTGGGATGGTCCGTTTGTCTGTG	3480
Db	2341		3480	Db	3421		3480

```
Db 3421 GTCCGAGTAGACATCCAGAGCAATATTTCCCTCGAGTGGATGGTTCCTCGTTGCTGTG 3480
Qy 3481 GATCGTAGTGTGATCGTCCGCTAGCTGAGGTAAACCTGGCGAATTAGTGTATCGCGGAC 3540
Db 3481 GATCGTAGTGTGATCGTCCGCTAGCTGAGGTAAACCTGGCGAATTAGTGTATCGCGGAC 3540
Qy 3541 GATAATACAGCCATTGAGATAACTGTCAGGTGATGGACAGTTCATTCCTCTTTCCGGGC 3600
Db 3541 GATAATACAGCCATTGAGATAACTGTCAGGTGATGGACAGTTCATTCCTCTTTCCGGGC 3600
Qy 3601 CTTCAAGGTGACACATATTGAGAGATAAGTATATATAAAGGTCCTTTTGTGCGAATGT 3660
Db 3601 CTTCAAGGTGACACATATTGAGAGATAAGTATATATAAAGGTCCTTTTGTGCGAATGT 3660
Qy 3661 ACAATACTTCGCAAAATCCCTTGATC 3687
Db 3661 ACAATACTTCGCAAAATCCCTTGATC 3687

RESULT 5
LOCUS I26124 6911 bp DNA linear PAT 07-OCT-1996
DEFINITION Sequence 4 from patent US 5556776.
ACCESSION I26124
VERSION I26124.1 GI:1605994
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 6911)
AUTHORS Tsuchiya, M. and Miwa, K.
TITLE Sucrose gene derived from coryneform bacteria
JOURNAL Patent: US 5556776-A 4 17-SEP-1996;
FEATURES
Location/Qualifiers
source 1..6911
/organism="unknown"
BASE COUNT 1575 a 1784 c 1892 g 1660 t
ORIGIN

Query Match 61.8%; Score 3687; DB 6; Length 6911;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTCGCTGACGCGCACCATTTGATGTGTGTCTACCGAGCTTGGGAGGCTTCTACATCT 60
Db 1 AGTCGCTGACGCGCACCATTTGATGTGTGTGTCTACCGAGCTTGGGAGGCTTCTACATCT 60
Qy 61 ACGCTCCGCTGCGCGTGGAGTGGGTGATTAACGGTGGGATCACGCGCGTGAAGTTGCG 120
Db 61 ACGCTCCGCTGCGCGTGGAGTGGGTGATTAACGGTGGGATCACGCGCGTGAAGTTGCG 120
Qy 121 GAACCCATGTGTCTCTGCTGAGTGGAGTGGGTGAGTGGGTGAGTGGGTGAGTGGGTGAGT 180
Db 121 GAACCCATGTGTCTCTGCTGAGTGGAGTGGGTGAGTGGGTGAGTGGGTGAGTGGGTGAGT 180
Qy 181 TCTGCAATTTTAAAGTTATGATCATCATCAGCTTGAAGGCTGAGTGAATTCAGTACGCTG 240
Db 181 TCTGCAATTTTAAAGTTATGATCATCATCAGCTTGAAGGCTGAGTGAATTCAGTACGCTG 240
Qy 241 CAACAGCAGCCCTCAAGTCCGAGATTAATTAACCTAGATCCGCTAGACATCAACATCATATA 300
Db 241 CAACAGCAGCCCTCAAGTCCGAGATTAATTAACCTAGATCCGCTAGACATCAACATCATATA 300
Qy 301 CGTCTATGTTGCTGGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTGCAAT 360
Db 301 CGTCTATGTTGCTGGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTGCAAT 360
Qy 361 ATCAAGAAATGAGTCTAGCAGTAAAGAAATGAGGAGAAATGAGGAGAAATGAGGAGAAATG 420
Db 361 ATCAAGAAATGAGTCTAGCAGTAAAGAAATGAGGAGAAATGAGGAGAAATGAGGAGAAATG 420
Qy 421 GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATCGGAACTCTCTGGAGAAC 480
Db 421 GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATCGGAACTCTCTGGAGAAC 480
Qy 481 CAGCACCTTAAAGAGCAGGATTCACCCCGCAACTCCCGACGATTTGTTCCCGGTTTATTTG 540
Db 481 CAGCACCTTAAAGAGCAGGATTCACCCCGCAACTCCCGACGATTTGTTCCCGGTTTATTTG 540
Qy 541 ATCTTCATAATCAGCGTGGAAACGCTGGCGGTTTCTTACGGGAACGACGAGCAGCGGA 600
Db 541 ATCTTCATAATCAGCGTGGAAACGCTGGCGGTTTCTTACGGGAACGACGAGCAGCGGA 600
Qy 601 GGAACACCGCGCAGTATCACCGCGAAACATGCGACGACGCTGATGTTGCCAAGCATGTTT 660
Db 601 GGAACACCGCGCAGTATCACCGCGAAACATGCGACGACGCTGATGTTGCCAAGCATGTTT 660
Qy 661 CGGCGCGGCTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTCTGTGTGAAGGG 720
Db 661 CGGCGCGGCTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTCTGTGTGAAGGG 720
Qy 721 TCCTGCTGTGGGCAATTCACCTCGAGGSCCTTTTCATCAACGCGATCGCTGTGTGCTC 780
Db 721 TCCTGCTGTGGGCAATTCACCTCGAGGSCCTTTTCATCAACGCGATCGCTGTGTGCTC 780
Qy 781 AAAACCCGATTTCAATTTTCCCGGCAACCCAGAGATCTTTGCCCGGCTGATCCATCGG 840
Db 781 AAAACCCGATTTCAATTTTCCCGGCAACCCAGAGATCTTTGCCCGGCTGATCCATCGG 840
Qy 841 GAAAAGTTGGATCAATCGATCACAGTAGCGCGGAAACTGACAATCTTTCTGAGCTTC 900
Db 841 GAAAAGTTGGATCAATCGATCACAGTAGCGCGGAAACTGACAATCTTTCTGAGCTTC 900
Qy 901 TCGATCTCTGCGAGCGCACCATCATCTTCTTCCGGGCACTGATGATGATGATGATGATG 960
Db 901 TCGATCTCTGCGAGCGCACCATCATCTTCTTCCGGGCACTGATGATGATGATGATGATG 960
Qy 961 ATACCACATACCAGCGCAATTCCTTGGCTAAAGAGAAATGTCAGGCTACGCGTACGCTAC 1020
Db 961 ATACCACATACCAGCGCAATTCCTTGGCTAAAGAGAAATGTCAGGCTACGCGTACGCTAC 1020
Qy 1021 ATTTGTTCAATTCGATGCTCCGCTGATCATATAGGGTTCGCGGAGCGTGGCGCTTTGCG 1080
Db 1021 ATTTGTTCAATTCGATGCTCCGCTGATCATATAGGGTTCGCGGAGCGTGGCGCTTTGCG 1080
Qy 1081 TTGCTGCGGCACTGCGCGGAGCAGATGTTGATGTTGATGCTGCGGAGCGGCTGATGTTG 1140
Db 1081 TTGCTGCGGCACTGCGCGGAGCAGATGTTGATGTTGATGCTGCGGAGCGGCTGATGTTG 1140
Qy 1141 CCGATGGAACGCTGATCTAGCTCTGTTCCAAACGCTTTTTCATCACGAGACGCCATGG 1200
Db 1141 CCGATGGAACGCTGATCTAGCTCTGTTCCAAACGCTTTTTCATCACGAGACGCCATGG 1200
Qy 1201 AAGCGCGGGAATCCAGACGCTGAGTACATTTTGGGCGTTTGAAGCTCACCGTCAACCG 1260
Db 1201 AAGCGCGGGAATCCAGACGCTGAGTACATTTTGGGCGTTTGAAGCTCACCGTCAACCG 1260
Qy 1261 ATGGAGTCCGCGCTGCGCGATGCGCGGCGCATGCGCGGCGGCGACGACACATAGCGA 1320
Db 1261 ATGGAGTCCGCGCTGCGCGATGCGCGGCGCATGCGCGGCGGCGACGACACATAGCGA 1320
Qy 1321 GTCAGTTGCTGACACGCTGCGAGGGGTATGAGCGTTATCGAGCGGCGACCGTCCACACCT 1380
Db 1321 GTCAGTTGCTGACACGCTGCGAGGGGTATGAGCGTTATCGAGCGGCGACCGTCCACACCT 1380
Qy 1381 CAACCGTCCGCGCTAAAATTTCTCGCTTGGCGATTCAGAAATCGCTAAAATCCAACCTG 1440
Db 1381 CAACCGTCCGCGCTAAAATTTCTCGCTTGGCGATTCAGAAATCGCTAAAATCCAACCTG 1440
Qy 1441 CAAATTTGCTGCTTTGACTCAACGCGCAGGTGCAAAAGTCCATTTAGGTGATCAAG 1500
Db 1441 CAAATTTGCTGCTTTGACTCAACGCGCAGGTGCAAAAGTCCATTTAGGTGATCAAG 1500
Qy 1501 TACTTTAAGTACGAGTAAAATCTATCCTGATTTTAAAGGAGTCCACCATGGAATCACTA 1560
Db 1501 TACTTTAAGTACGAGTAAAATCTATCCTGATTTTAAAGGAGTCCACCATGGAATCACTA 1560
```

Db 1501 TACTTTAAGTACGAGTAAACTATCCTGATTTTAAAGGAGTCCCAACCATGGAATCACTA 1560  
Qy 1561 TCTGCAAGACGAGCAAGAGTGGGCAAGACAGTTGAGTCCCTAAATCGCACCCCTTGGCCA 1520  
Db 1561 TCTGCAAGACGAGCAAGAGTGGGCAAGACAGTTGAGTCCCTAAATCGCACCCCTTGGCCA 1520  
Qy 1621 ACAGGGTGGAACTTGGGGCTTGCACACAGGATCCTCACACATGAGTACCTACCAAGAGC 1680  
Db 1621 ACAGGGTGGAACTTGGGGCTTGCACACAGGATCCTCACACATGAGTACCTACCAAGAGC 1680  
Qy 1681 TCATTCCGATGATGAGCTGGGAAGTGTCAATCAAGAACTGCAAGGCAATCTTGTGG 1740  
Db 1681 TCATTCCGATGATGAGCTGGGAAGTGTCAATCAAGAACTGCAAGGCAATCTTGTGG 1740  
Qy 1741 ATGAATACGTGGGACTAACCGGTGACGATGAACAGCTACTTTAAACCAATTCGCAAG 1800  
Db 1741 ATGAATACGTGGGACTAACCGGTGACGATGAACAGCTACTTTAAACCAATTCGCAAG 1800  
Qy 1801 AGTTCACTGACCACTACGATCGTGTGATGAAGAGTCTACAGCCAGATGGTGCAACC 1860  
Db 1801 AGTTCACTGACCACTACGATCGTGTGATGAAGAGTCTACAGCCAGATGGTGCAACC 1860  
Qy 1861 CTGATCCATAGCAAGACGTGCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAAG 1920  
Db 1861 CTGATCCATAGCAAGACGTGCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAAG 1920  
Qy 1921 TTCAAACTCTTGGCATCGGGGAAAGGACATCGCTTTCATTGAACCATCTTCTCT 1980  
Db 1921 TTCAAACTCTTGGCATCGGGGAAAGGACATCGCTTTCATTGAACCATCTTCTCT 1980  
Qy 1981 GTGAGGACTGCAAAAGTGCAGGGCTGCACCTTAAACCTGTGGAGCAACGCTCGATT 2040  
Db 1981 GTGAGGACTGCAAAAGTGCAGGGCTGCACCTTAAACCTGTGGAGCAACGCTCGATT 2040  
Qy 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGGTTTGGGCATTTGTC 2100  
Db 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGGTTTGGGCATTTGTC 2100  
Qy 2101 CCGCGCAAAACATCGTGTGGCACTGTGGAAGAAAGCGACGCCATCCGCGG 2160  
Db 2101 CCGCGCAAAACATCGTGTGGCACTGTGGAAGAAAGCGACGCCATCCGCGG 2160  
Qy 2161 AACTGTGGAAGGCCAGTACTGCTTCTCCAGSTTCCATCCTGTAGTGCACAACAT 2220  
Db 2161 AACTGTGGAAGGCCAGTACTGCTTCTCCAGSTTCCATCCTGTAGTGCACAACAT 2220  
Qy 2221 GCCACCATCTGTTGATGAGCAGCAGTATCCAGCTGGAACCGCTGATCACTACCG 2280  
Db 2221 GCCACCATCTGTTGATGAGCAGCAGTATCCAGCTGGAACCGCTGATCACTACCG 2280  
Qy 2281 TCTCATGAGCAATTAAGCTGGCTAGAAACAAAAGAAAGTACTGTGTGGGGTATG 2340  
Db 2281 TCTCATGAGCAATTAAGCTGGCTAGAAACAAAAGAAAGTACTGTGTGGGGTATG 2340  
Qy 2341 CACACAGACTTCCAGTTTGGCCCTCGCTACCATGTGACTCCTCCGAGGCGAGGCTC 2400  
Db 2341 CACACAGACTTCCAGTTTGGCCCTCGCTACCATGTGACTCCTCCGAGGCGAGGCTC 2400  
Qy 2401 AATGATCCCAACGAAATGACGTGGAGTACCCCTCCAGCTACTACTACCAAGCAGAT 2460  
Db 2401 AATGATCCCAACGAAATGACGTGGAGTACCCCTCCAGCTACTACTACCAAGCAGAT 2460  
Qy 2461 CCAGGTTCCCTTCGCACCAAGCGACCGGTGGGCTCACACACCGCCGCTTGACC 2520  
Db 2461 CCAGGTTCCCTTCGCACCAAGCGACCGGTGGGCTCACACACCGCCGCTTGACC 2520  
Qy 2521 GGACCGGAGGATTCAGTGGACGACCTCGCCGACGCTCTTACCCGGATGATCCTAT 2580  
Db 2521 GGACCGGAGGATTCAGTGGACGACCTCGCCGACGCTCTTACCCGGATGATCCTAT 2580  
Qy 2581 GACCTGATGATGCTATTCGGTGGAGCGGTATTTTACTGACGGCACACTTAACTTTTC 2640  
Db 2581 GACCTGATGATGCTATTCGGTGGAGCGGTATTTTACTGACGGCACACTTAACTTTTC 2640

Qy 2641 TACACGGCAACCTTAAATAATTCACGAAAGCCCGCGCACCAAAACCTTGTGGAAGTC 2700  
Db 2641 TACACGGCAACCTTAAATAATTCACGAAAGCCCGCGCACCAAAACCTTGTGGAAGTC 2700  
Qy 2701 GAGGACCCAACTGGGCTGATGGCGGCATTCATCGCCTTCCCTTAAATAATCCGCTTATC 2760  
Db 2701 GAGGACCCAACTGGGCTGATGGCGGCATTCATCGCCTTCCCTTAAATAATCCGCTTATC 2760  
Qy 2761 GAGGACCCCGCAGCGGTTTACACCCATTTACCGGATCCCATGATCAGCCCTGATGGT 2820  
Db 2761 GAGGACCCCGCAGCGGTTTACACCCATTTACCGGATCCCATGATCAGCCCTGATGGT 2820  
Qy 2821 GATGTTGAAACATGGTTCCTGGGGCCCAACCGGAAAAACCTCACCGGTGCACGCTTCTA 2880  
Db 2821 GATGTTGAAACATGGTTCCTGGGGCCCAACCGGAAAAACCTCACCGGTGCACGCTTCTA 2880  
Qy 2881 TACCGTTCACAGATCTTGAACCTGGGAATTTCTCCGGTGAATCACCTTTGACCTCAGT 2940  
Db 2881 TACCGTTCACAGATCTTGAACCTGGGAATTTCTCCGGTGAATCACCTTTGACCTCAGT 2940  
Qy 2941 GATGCACAACCTGGTTCCTCTGATCTCGTTCCTCCGATGGCTACATGCGGAATGCCCC 3000  
Db 2941 GATGCACAACCTGGTTCCTCTGATCTCGTTCCTCCGATGGCTACATGCGGAATGCCCC 3000  
Qy 3001 AACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGAGCTGCTGATTTCTGT 3060  
Db 3001 AACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGAGCTGCTGATTTCTGT 3060  
Qy 3061 CCACAGGATTTGGACCGAATCCACGATGAGTTACTCTAGCGAAGCTCTGACCACTGTC 3120  
Db 3061 CCACAGGATTTGGACCGAATCCACGATGAGTTACTCTAGCGAAGCTCTGACCACTGTC 3120  
Qy 3121 GGATATGCTCGACAGCTTGAAGAACGAGCCTTCCCGCTTCCGAGGATTCAGCGAG 3180  
Db 3121 GGATATGCTCGACAGCTTGAAGAACGAGCCTTCCCGCTTCCGAGGATTCAGCGAG 3180  
Qy 3181 CTGGATTTCCGCCATGAATTTACGACCGCAGGTTCAGTAAACGGTTCTGATSCCTGG 3240  
Db 3181 CTGGATTTCCGCCATGAATTTACGACCGCAGGTTCAGTAAACGGTTCTGATSCCTGG 3240  
Qy 3241 CTCGTGGGTGATGGGGTCCCGCGCAGGATGATACCCAACAGTTGTCACAGGAAGA 3300  
Db 3241 CTCGTGGGTGATGGGGTCCCGCGCAGGATGATACCCAACAGTTGTCACAGGAAGA 3300  
Qy 3301 TGGGTGCACTGCTGACTGTGCCCGCAGCTTCAATTTGCCAACCCACCGATCTACCAA 3360  
Db 3301 TGGGTGCACTGCTGACTGTGCCCGCAGCTTCAATTTGCCAACCCACCGATCTACCAA 3360  
Qy 3361 GAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATTAGTTCTGAACCT 3420  
Db 3361 GAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATTAGTTCTGAACCT 3420  
Qy 3421 GTCCGAGTAGACATCCGAGGCAATTTTCCCTCGAGTGGAGTGTCCGTTTGTCTGTG 3480  
Db 3421 GTCCGAGTAGACATCCGAGGCAATTTTCCCTCGAGTGGAGTGTCCGTTTGTCTGTG 3480  
Qy 3481 GATCTGTAGTGTGCTGCTGCTAGCTGAGSTAAACCTGCGCAATTAGTATCAGCGGAC 3540  
Db 3481 GATCTGTAGTGTGCTGCTGCTAGCTGAGSTAAACCTGCGCAATTAGTATCAGCGGAC 3540  
Qy 3541 GATAATACAGCCATTGAGATTAATCTCAGGTGATGACAGGTTTCATTCGCTTTCCGGGC 3600  
Db 3541 GATAATACAGCCATTGAGATTAATCTCAGGTGATGACAGGTTTCATTCGCTTTCCGGGC 3600  
Qy 3601 CTTCAAAGGTGACACTATTGAGAGATACTAATAAGGGTCTTTTGTGCGCAATTTGT 3660  
Db 3601 CTTCAAAGGTGACACTATTGAGAGATACTAATAAGGGTCTTTTGTGCGCAATTTGT 3660  
Qy 3661 ACAATATCTCGCAAAATCCCTTGATC 3687  
Db 3661 ACAATATCTCGCAAAATCCCTTGATC 3687

RESULT 6  
AX122988  
LOCUS AX122988 1983 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 2904 from Patent EP1108790.  
ACCESSION AX122988  
VERSION AX122988.1 GI:14040476  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Corynebacteriidae; Corynebacterium.  
REFERENCE  
1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
Novel polynucleotides  
Patent: EP 1108790-A 2904 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
Location/Qualifiers  
1..1983  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
BASE COUNT 410 a 508 c 558 g 507 t  
ORIGIN  
Query Match 31.4%; Score 1874.2; DB 6; Length 1983;  
Best Local Similarity 96.6%; Pred. No. 0;  
Matches 1915; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
Qy 3779 ATGACCATAGGACCTCGCGCAACGCATCCTCGCGACATTTGGCGCGCAAGACACATTT 3838  
Db 1 ATGACCATAGGACCTCGCGCAACGCATCCTCGCGACATTTGGCGCGCAAGACACATTT 60  
Qy 3839 GTCCGCCCGCACACTGTGCAACCGCTTACGCTCGTCTCAAGACACACCAAGGATGTG 3898  
Db 61 GTCCGCCCGCACACTGTGCAACCGCTTACGCTCGTCTCAAGACACCAAGGATGTG 120  
Qy 3899 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAACTGGCGGCAATG 3958  
Db 121 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAAGCGGTGATG 180  
Qy 3959 TTCCAGATATCTGCGGGCAGCGGATGTGATCATGTTTCAAGAACTCGATGACGCA 4018  
Db 181 TTCCAGATATCTGCGGGCAGCGGATGTGATCATGTTTCAAGAACTCGATGACGCA 240  
Qy 4019 ACCTCCAAAGACATCGGTGTGTCACAGAGAGCTCAAGAGATTTGTGCTTAACACGCC 4078  
Db 241 ACCTCCAAAGACATCGGTGTGTCACAGAGAGCTCAAGAGATTTGTGCTTAACACGCC 300  
Qy 4079 AACTGGTTACGGCTGCTGTAAGGTATTGGCGGACATTTTCGTCGCGCTGATTTCCAATC 4138  
Db 301 AACTGGTTACGGCTGCTGTAAGGTATTGGCGGACATTTTCGTCGCGCTGATTTCCAATC 360  
Qy 4139 TTGGTTGGTGGCGGTCTGCTCATGCTATCAACAATGTGTTGGTTCGCGAGGATCTGTTTC 4198  
Db 361 TTGGTTGGTGGCGGTCTGCTCATGCTATCAACAATGTGTTGGTTCGCGAGGATCTGTTTC 420  
Qy 4199 GTTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGCTCGATGATCAAC 4258  
Db 421 GTTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGCTCGATGATCAAC 480  
Qy 4259 CTGATGCGATCTCGCGGTTCGGTCTTTCGCCAGTGTGTTGGTTTCAACCGCAACCAAG 4318  
Db 481 CTGATGCGATCTCGCGGTTCGGTCTTTCGCCAGTGTGTTGGTTTCAACCGCAACCAAG 540  
Qy 4319 CGTTTCGGTGGCAATGATGTTCTTGGCGCGGCTGTTGGTATGCGGATGTTTCCCAACC 4378  
Db 541 CGTTTCGGTGGCAATGATGTTCTTGGCGCGGCTGTTGGTATGCGGATGTTTCCCGAGC 600  
Qy 4379 CTGGTTAACGGCTACGACGTGGCGGCCACCATGACCGCGGCGGCAATGCAATGTGCTGC 4438  
Db 601 TTGGTGAACGGCTACGACGTGGCGGCCACCATGACCGCGGCGGCAATGCAATGTGCTGC 660

Qy 4439 CTGTTTGGTTGGATGTGCTCAAGCTGGTTACACAGGACACCGTGCTTCTGTGCTGGTG 4498  
Db 661 CTGTTTGGTTAGATGTTGCCAAGCCGTTACACAGGACCGTGCTTCTGTGCTGGTG 720  
Qy 4499 GTCTCTTGATCTGCGCAACGATCGAGAGTTCCTGTCACAGGACGACTCATGGGACGCTCA 4558  
Db 721 GTTCTTGGATTTGCGCAACGATCGAGAGTTCCTGTCACAGGACGACTCATGGGACGCTCA 780  
Qy 4559 GACTTCTCTGATCACCCTCAGTGTGACTCTGCTGCTTCAACGGCTTCTTACGTTTACGTT 4618  
Db 781 GACTTCTCTGATCACCCTCAGTGTGACTCTGCTGCTTCAACGGATCTTACATTCATCGCC 840  
Qy 4619 ATTGCTCCAGCAATCGGCTGGTGGTGCATGCTTGGCGACAGCTGTCGAGGAGCTCAT 4678  
Db 841 ATTGCTCCAGCAATCGGCTGGTGGTGCATGCTTGGCGACAGCTGTCGAGGAGCTTAT 900  
Qy 4679 GATTTCGGTGTGCTCAGTGGGCTGCTGCTTTCGGTCTGCTTACTCACCAGTGTATC 4738  
Db 901 GATTTCGGTGTGCTCAGTGGGCTGCTGCTTTCGGTCTGCTTACTCACCAGTGTATC 960  
Qy 4739 ACTGCTGTGACCCAGTCTTCCGCAATTTGAGCTGGAGCTGTTCAACAGGCTGGATCC 4798  
Db 961 ACTGCTGTGACCCAGTCTTCCGCAATTTGAGCTGGAGCTGTTCAACAGGCTGGATCC 1020  
Qy 4799 TTCACTCTCGCAACCGCATCCATGCGCAATATCGCGCAGGCTGCGAGCTGTTTGGCAGTG 4858  
Db 1021 TTCACTCTCGCAACCGCATCTATGGCTATATCGCGCAGGCTGCGAGCTGTTTGGCAGTG 1080  
Qy 4859 TTCTTCTTAGCGAAGAGTGAAGCTCAAGGCGCTTGCAGGTGCTTCAAGTGTCTCCGCT 4918  
Db 1081 TTCTTCTTAGCGAAGAGTGAAGCTCAAGGCGCTTGCAGGTGCTTCAAGTGTCTCCGCT 1140  
Qy 4919 GTTCTTGTGTTTACAGAGCTGCGATCTTTCGGTGTGAACCTTTCGCTGCGCTGGCGCTTC 4978  
Db 1141 GTTCTTGTGTTTACAGAGCTGCGATCTTTCGGTGTGAACCTTTCGCTGCGCTGGCGCTTC 1200  
Qy 4979 TACATTTGATTCGGTACCGCAGCTATTCGGTGGCGCTTTCGATGCTGCTTTCGATATCAAG 5038  
Db 1201 TACATTTGATTCGGTACCGCAGCTATTCGGTGGCGCTTTCGATGCTGCTTTCGATATCAAG 1260  
Qy 5039 GCAGTTGGTTCGGCGCTGCGAGTTCCTTGGTGTGTTTCTTATGATGCTGCCAGATG 5098  
Db 1261 GCAGTTGGTTCGGCGCTGCGAGTTCCTTGGTGTGTTTCTTATGATGCTGCCAGATG 1320  
Qy 5099 GTCATGTTCTTGGTTGGCGGTAGTTACCTTGTTCATCGCATTCGGCGCAGCGATGCT 5158  
Db 1321 GTCATGTTCTTGGTTGGCGGTAGTTACCTTGTTCATCGCATTCGGCGCAGCGATGCT 1380  
Qy 5159 TATGGCTTTACTTGGTTGGCGGTAGTTACCTTGTTCATCGCATTCGGCGCAGCGATGCT 5218  
Db 1381 TATGGCTTTACTTGGTTGGCGGTAGTTACCTTGTTCATCGCATTCGGCGCAGCGATGCT 1440  
Qy 5219 GTGCTGCGAGAACGACCAAGCGAGAGCAGGACCGCGAGATTTTCAAGAGTATCC 5278  
Db 1441 GTGCTGCGAGAACGACCAAGCGAGAGCAGGACCGCGAGATTTTCAAGAGTATCC 1500  
Qy 5279 ACCATCATCCAGGACCTTTTACCGGTGAAGCTATCGCACTGAGCAGCGTTCAGCGATGCC 5338  
Db 1501 ACCATCATCCAGGACCTTTTACCGGTGAAGCTATCGCACTGAGCAGCGTTCAGCGATGCC 1560  
Qy 5339 ATGTTTGGCAGCGGAAAGCTTGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5398  
Db 1561 ATGTTTGGCAGCGGAAAGCTTGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1620  
Qy 5399 GTTTCACAGTACGCGGAAAGCTGCTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5458  
Db 1621 GTTTCACAGTACGCGGAAAGCTGCTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1680  
Qy 5459 CGCATTAAGGCTGAGGATGTTTCAATGTGGATATCTTGTGATGCACTGTTGTTGCAACACC 5518  
Db 1681 CGCATTAAGGCTGAGGATGTTTCAATGTGGATATCTTGTGATGCACTGTTGTTGCAACACA 1740

```
Qy 5519 GTAAACCTCAACGGCAGGACCTTTAAACCGCTGAAGAAAGCAGGGCGATGAAGTCAAAGCA 5578
Db 1741 GTAAACCTCAACGGCAGGACCTTTAAACCGCTGAAGAAAGCAGGGCGATGAAGTCAAAGCA 1800
Qy 5579 GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGACGGTTATGAGGTAACC 5638
Db 1801 GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGACGGTTATGAGGTAACC 1860
Qy 5639 ACGCCGATGTTGTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTACGGTTTGGGC 5698
Db 1861 ACGCCGATGTTGTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTACGGTTTGGGC 1920
Qy 5699 GAAATGAAGCGGAGCGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 5758
Db 1921 GAAATGAAGCGGAGCGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 1980
Qy 5759 CCA 5761
Db 1981 CCA 1983

RESULT 7
BD165105
LOCUS Novel polynucleotide. 1983 bp DNA linear PAT 17-JAN-2003
DEFINITION Novel polynucleotide.
ACCESSION BD165105
VERSION BD165105.1 GI:27870917
KEYWORDS JP 2002191370-A/2904.
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1983)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoo,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2904 09-JUL-2002;
KYOWA HAKKO KOGYO CO LTD
OS Corynebacterium glutamicum
PN JP 2002191370-A/2904
PD 09-JUL-2002
PF 15-DEC-2000 JP 2000405096
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,
PI KEIKO OCHIAI,
PI HARUHIKO YOKOI, NAKKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO
PI OZAKI
PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC
C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/
PC 04, C12P13/08,
PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC
G01N33/566,
PC G01N33/569, G01N33/68, G01N37/00, C12P21/08, (C12N1/21, C12P1:15),
PC (C12N1/21, C12P1:13), (C12N1/21, C12P1:01), (C12P13/08, C12P1:15),
PC C12N15/00,
PC C12N5/00, C12N15/00
CC Novel polynucleotide
FH Key Location/Qualifiers
FT Source 1..1983
FT /organism="Corynebacterium glutamicum".
FEATURES
Location/Qualifiers
1..1983
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 410 a 508 c 558 g 507 t
ORIGIN
Query Match 31.4%; Score 1874.2; DB 6; Length 1983;
Best Local Similarity 96.6%; Pred. No. 0;
Matches 1915; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
Qy 3779 ATGGACCATATAGGACCTCGCGCAACGCGATCTCGCGACATGTGGCGGAAGACAACATT 3838
|||||
```

```
Db 1 ATGGACCATATAGGACCTCGCGCAACGCGATCTCGCGACATTTGGCGGAAGACAACATT 60
Qy 3839 GTGCGCGCGCACACTGTGCAACCGGTTTACCGCTCGTGTCTCAAAGACACCAAGAGTGTG 3898
Db 61 GTGCGCGCGCACACTGTGCAACCGGTTTACCGCTCGTGTCTCAAAGACACCAAGAGTGTG 120
Qy 3899 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCACTTTGAAACTGGGGGCATG 3958
Db 121 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCACTTTGAAACTGGGGTGTATG 180
Qy 3959 TTCCAGATCATGTCGGCCAGGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCA 4018
Db 181 TTCCAGATCATGTCGGCCAGGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCA 240
Qy 4019 ACCTCCAAAGACATCGCTGTGTCACAGACAGAGCTCAAAGATGTTGTGGCTTAAACAACGCC 4078
Db 241 ACCTCCAAAGACATCGCTGTGTCACAGACAGAGCTCAAAGATGTTGTGGCTTAAACAACGCC 300
Qy 4079 AACTGTTTCAGCCGTCGTGTAAGGTATTGGGGGACATTTTCGTCGCCGCTGATCCAATC 4138
Db 301 AACTGTTTCAGCCGTCGTGTAAGGTATTGGGGGACATTTTCGTCGCCGCTGATCCAATC 360
Qy 4139 TTGGTTGTGGGGTCTGCTCATGGCTATCAACAATGTTGGTGGCCAGATCTGTTC 4198
Db 361 TTGGTTGTGGGGTCTGCTCATGGCTATCAACAATGTTGGTGGCCAGATCTGTTC 420
Qy 4199 GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258
Db 421 GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 480
Qy 4259 CTGATGGCATCTGCCCGCTTCGCTTCATGGCTATCAACAATGTTGGTGGCCAGACCAAG 4318
Db 481 CTGATGGCATCTGCCCGCTTCGCTTCATGGCTATCAACAATGTTGGTGGCCAGACCAAG 540
Qy 4319 CGTTTCGGTGGCAATGAGTTCTCGGCGCGGCATTTGTTATGGCATGTTGTTCCCAACC 4378
Db 541 CGTTTCGGTGGCAATGAGTTCTCGGCGCGGCATTTGTTATGGCATGTTGTTCCCGAGC 600
Qy 4379 CTGGTTAAGCGGTACGACGTTGGCCGCCACCATGACCGGGGCGAATGCCATGTTGGTCC 4438
Db 601 TTGGTGAACGGGTACGACGTTGGCCGCCACCATGTTGGTGGCGGAATGCCATGTTGGTCC 560
Qy 4439 CTGTTTGGTTTGGATGTTGTTCTAAAGCTGTTTACCAGGCGACCGTCTCTCTGTTGTTG 4498
Db 661 CTGTTTGGTTTGGATGTTGTTCTAAAGCTGTTTACCAGGCGACCGTCTCTCTGTTGTTG 720
Qy 4499 GTCTCTGGATCTTGGCAACGATCGAGAGTTCCTGCACAAGGACTCATGGGCACTGCA 4558
Db 721 GTTCTCTGGATCTTGGCAACGATCGAGAGTTCCTGCACAAGGACTCATGGGCACTGCA 780
Qy 4559 GACTTCTGATCAACCCAGTGTGTTGACTCTGCTGCTCACCAGGCTTCCTTACGTTCAATTGCT 4618
Db 781 GACTTCTGATCACTCCAGTGTGCTGCTGCTCACCAGGATTCCTTACATTCATCGCC 840
Qy 4619 ATTGGTCCACCAATGCGCTGGGTGGGTGACTTGTCTGGCACACGGTCTGAGGAGACTCTAT 4678
Db 841 ATTGGTCCACCAATGCGCTGGGTGGGTGACTTGTCTGGCACACGGTCTGAGGAGACTCTAT 900
Qy 4679 GATTTCGGTGGTCCAGTCCGCGGCTGCTTTTCGGTCTGCTTACTACCAATCGTATC 4738
Db 901 GATTTCGGTGGTCCAGTCCGCGGCTGCTTCTCGGTCTGCTCTGCTCTACCAATCGTATC 960
Qy 4739 ACTGGTCTGCACCAGTCCCTTCCCGCAATTGAGCTGAGCTGTTCAACACCGGTGGATCC 4798
Db 961 ACTGGTCTGCACCAGTCCCTTCCCGCAATTGAGCTGAGCTGTTTAAACAGGGTGGATCC 1020
Qy 4799 TTCACTTTCGCAACCGCATCCATGGCCAATATCCGCGAGGGTGCACATGTTTGGCAGTG 4858
Db 1021 TTCACTTTCGCAACCGCATCTATGGTAATATCCGCGAGGGTGCAGATGTTTGGCAGTG 1080
Qy 4859 TTCTTCTACGAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGTTCTCCGCT 4918
Db 1081 TTCTTCTGCGGAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGTTCTCCGCT 1140
```

QY 4919 GTCTGGTATTACAGAGCTCGATCTCGGTGTGAACCTTCGCTGGCTGGCGGTC 4978  
DB 1141 GTCTGGTATTACAGAGCTCGATCTCGGTGTGAACCTTCGCTGGCTGGCGGTC 1200  
QY 4979 TACATTGGTATCGGACGCTATCGGTGGCGCTTTGATGTCACCTTTTGAATCAAG 5038  
DB 1201 TACATTGGTATCGGACGCTATCGGTGGCGCTTTGATGTCACCTTTTGAATCAAG 1260  
QY 5039 GCAGTTGGCTGGCGCTCGAGTTCTTGGGTGTGTTCTTATTGATGCTCCAGATG 5098  
DB 1261 GCAGTTGGCTGGCGCTCGAGTTCTTGGGTGTGTTCTTATTGATGCTCCAGATG 1320  
QY 5099 GTCAATGTTCTGGTTTGGCGGTAGTACCTTTGTATCGCATTCGGCGCAGCATTC 5158  
DB 1321 GTCAATGTTCTGGTTTGGCGGTAGTACCTTTGTATCGCATTCGGCGCAGCATTC 1380  
QY 5159 TATGGCTTTACTTGGTTTCGCGCAACGCGCAGCATTCAGATGTCACCGCTCTCCA 5218  
DB 1381 TATGGCTTTACTTGGTTTCGCGCAACGCGCAGCATTCAGATGTCACCGCTCTCCA 1440  
QY 5219 GTGCTTCAGGACGACCAAGCGGAGAGCAGCAGCGGCGAGATTTTCAACGATTC 5278  
DB 1441 GTGCTTCAGGACGACCAAGCGGAGAGCAGCAGCGGCGAGATTTTCAACGATTC 1500  
QY 5279 ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCACTGACGAGCTCAGCATGCC 5338  
DB 1501 ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCACTGACGAGCTCAGCATGCC 1560  
QY 5339 ATGTTTGCAGCGGAAAGCTTGGCTCAGGTGTGGCATCGTCCCAACGAGGCGAGTG 5398  
DB 1561 ATGTTTGCAGCGGAAAGCTTGGCTCAGGTGTGGCATCGTCCCAACGAGGCGAGTG 1620  
QY 5399 GTTTCACGAGTGAAGATCGGTGGCTTCCGATCTTGGTACGCTTCCGAGTC 5458  
DB 1621 GTTTCACGAGTGAAGATCGGTGGCTTCCGATCTTGGTACGCTTCCGAGTC 1680  
QY 5459 CGCACAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGACAC 5518  
DB 1681 CGCACAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGACAC 1740  
QY 5519 GTAACCTCAACGGCAGCAGCTTACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA 5578  
DB 1741 GTAACCTCAACGGCAGCAGCTTACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA 1800  
QY 5579 GGGAGGTGCTGTGAATTCGATTTGATGCCATTAAGCTGCGAGTTATGAGGTAACC 5638  
DB 1801 GGGAGGTGCTGTGAATTCGATTTGATGCCATTAAGCTGCGAGTTATGAGGTAACC 1860  
QY 5639 ACCCGATGTTTTCGATTTACAGAAACCGGACCTGTAACACATTTACGTTTGGCG 5698  
DB 1861 ACCCGATGTTTTCGATTTACAGAAACCGGACCTGTAACACATTTACGTTTGGCG 1920  
QY 5699 GAAATGAAGCGGAGCCACCTGCTCAAGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 5758  
DB 1921 GAAATGAAGCGGAGCCACCTGCTCAAGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 1980  
QY 5759 CCA 5761  
DB 1981 CCA 1983

RESULT 8  
LOCUS BD094213 1656 bp DNA linear PAT 27-AUG-2002  
DEFINITION Genes for heat resistant enzymes of amino acid biosynthetic pathway  
derived from thermophilic coryneform bacteria.  
BD094213  
ACCESSION BD094213  
VERSION BD094213.1 GI:22639801  
KEYWORDS WO 0125447-A/71.  
SOURCE Corynebacterium thermoaminogenes  
ORGANISM Corynebacterium thermoaminogenes  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE  
AUTHORS

TITLE  
JOURNAL

COMMENT

FEATURES  
source

BASE COUNT  
ORIGIN

Query Match 24.4%; Score 1457; DB 6; Length 1656;  
Best Local Similarity 93.2%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 110; Indels 2; Gaps 2;

QY 2031 ACGTCGATCTTCAACACCATCGAAGAGGTCCCAACCGCCGTCACCCAGGTTGG 2090  
DB 1 ACGCCGATCTTCAACACTATCGAAGAGGTCCCAACCGCCGTCACCCAGGTTGG 60  
QY 2091 GCATTTTCGCGCGCGCAAAACATCGTGTGGTGGCACTGTGAAGGAAAGCCGAC - 2149  
DB 61 GTACTTTTCGCGCGCGCAAAATATCGTGTGGTGGCACTGTGAAGGAAAGCCGAC 120  
QY 2150 GCCATCCGCGAAGTGTGGAAGGCCAGTGTCTTCTTCCAGAGTTCCATCTGTAG 2209  
DB 121 GCCATCCGCGAAGTGTGGAAGGCCAGTGTCTTCTTCCAGAGTTCCATCTGTAG 180  
QY 2210 ATGCACACATGCCACCATCATCTGTTGGATGAAGCAGCATTCACAGTGGAAAGCGT 2269  
DB 181 ATGCACACACAGCCACCATCATCTGTTGATGAAGCAGCATTCACAGTGGAAAGTGT 240  
QY 2270 GATCACTACCTCTCATCGAGCAATTAAGCTGGCTAGAACAAAGGAAAGTACTGT 2329  
DB 241 GACCATACCGTCTCATGGACAATTAAGCTGGCTAGAACAAAGGAAAGTACTGT 300  
QY 2330 GTGGGGCTATGCACACAGAACTTCCAGTTTCCGCCCTGCGTACCATGTGACTCTCGC 2389  
DB 301 GTGGGGCTATGCACACAGAACTTCCAGTTTCCGCCCTGCGTACCATGTGACTCTCGC 360  
QY 2390 AGGGAGGCTCAATCATCCCAACGGAATGTACGTGATGGAGATACCTCCACGCTACT 2449  
DB 361 AGGGAGGCTCAATCATCCCAACGGAATGTACGTGATGGAGATACCTCCACGCTACT 420  
QY 2450 ACCAGCAGATCCAGGTTTCCCTTCGCAACAAAGCGCAGCGGTGGGCTCACACCACA 2509  
DB 421 ACCAGCAGATCCAGGTTTCCCTTCGCAACAAAGCGCAGCGGTGGGCTCACACCACA 480

Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1 (bases 1 to 1656)  
Hirano, S., Nonaka, G., Matsuzaki, Y., Akiyoshi, N., Nakamura, K.,  
Kimura, E., Osumi, T., Matsui, K., Kawahara, Y., Kurahashi, O.,  
Nakamatsu, T. and Sugimoto, S.  
Genes for heat resistant enzymes of amino acid biosynthetic pathway  
derived from thermophilic coryneform bacteria  
Patent: WO 0125447-A 71 12-APR-2001;  
AJINOMOTO CO INC, SEIKO HIRANO, GEN NONAKA, YUMI MATSUZAKI, NAOKI  
AKIYOSHI, KANAE NAKAMURA, EIICHIRO KIMURA, TSUYOSHI OSUMI, KAZUHIKO  
MATSUI, YOSHIO KAWAHARA, OSAMU KURAHASHI, TSUYOSHI NAKAMATSU,  
SHINICHI SUGIMOTO  
OS Corynebacterium thermoaminogenes  
PN WO 0125447-A/71  
PD 12-APR-2001  
PF 04-OCT-2000 WO 2000JP006913  
PI 04-OCT-1999 JP 99P 282716, 01-NOV-1999 JP 99P 311147 PR  
21-APR-2000 JP 00P 120687  
PI SEIKO HIRANO, GEN NONAKA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE  
PI NAKAMURA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KAZUHIKO  
PI EIICHIRO KIMURA, TSUYOSHI OSUMI, KAZUHIKO  
MATSUI, YOSHIO KAWAHARA,  
PI OSAMU KURAHASHI, TSUYOSHI NAKAMATSU, SHINICHI SUGIMOTO PC  
C12N15/60, C12N15/54, C12N15/53, C12N15/31, C12N15/56, C12N9/88, PC  
C12N9/12,  
PC C12N9/04, C07K14/34, C12N9/26, C12N13/04  
CC  
FH Key Location/Qualifiers  
FT CDS (309)..(1595).

1. 1656  
/organism="Corynebacterium thermoaminogenes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:161484"

BASE COUNT 392 a 456 c 432 g 376 t

Qy	2510	CGCCGTTGACCGGACCCGACGATGTCAGTGGAGCGACCTGCGCGAGCGCTCTTTACCCGG	2569	Db	1561	C-TTTCGCGACCTTCAAAGGTGACACTATTGAGAGATAGTATATAAAAGGCGCTTCTG	1619
Db	481	CGCCGTTGACCGGACCCGACGGAATTCAGTGGAGCGACCTGCGCGATGCTTCTTTACCCGG	540	Qy	3650	TGGCGAATTTGTACAAATACTTCGCAAAATCCCTTGAT	3686
Qy	2570	ATGCATCTATGACCTGATGATGCTATTCGCGTGGAGCGGATTTACTGACGCGACAC	2629	Db	1620	TGGCGGATTTGACAAATACTTCGCAAAATCCCTTGAT	1656
Db	541	ATGTATCTATGACCTGATGATGCTATTCGCGGAGCGGATTTTCTGAGCGGACGC	600	RESULT 9			
Qy	2630	TTAAACTTTTCTACACCGCAACCTTAAATTTGACGAAAGCGCGCCACCCAAAC	2689	AX069134			
Db	601	TTAAACTTTTCTACACCGCAACCGAAATTTGACGCAAGCGCGCCACCCAAAC	660	LOCUS			
Qy	2690	TTGTGCAAGTCAGACCCCACTGGCTGATGSGCGCATTCATCGCGTTCGCTTAA	2749	DEFINITION			
Db	661	TCGTGCAAGTCAGACCCCACTGGCTGATGSGCGCATTCATCGCGTTCGCTTAA	720	ACCESSION			
Qy	2750	ATCCGCTTATCAGCAGGACCGCCAGCGGTTTTCACACCCCATTCACCGATCCCATGATCA	2809	VERSION			
Db	721	ATCCGCTTATCAGCAGGACCGCCAGCGGTTTTCACCGCCACTACCGCATCCCATGATCA	780	KEYWORDS			
Qy	2810	GCCTGATGGTGATGGTTGGAACATGGTTCTTGGGGGCCCAACGGGAAACCTACCGGTG	2869	SOURCE			
Db	781	GCCTGATGGGATGGTTGGAAGATGGTCTTGGGGCTCACGCGGAAACCTACCGGTG	840	ORGANISM			
Qy	2870	CACGCGTTCTATACCGCTCGACAGATCTGAAACTTGGAACTTCCTCGGTGAAATCACCT	2929	REFERENCE			
Db	841	CACGCGTTCTATACCGCTCGGAGATCTGAAACTGGGAATTCCTCGGTGAAATCACCT	900	AUTHORS			
Qy	2930	TTGACCTCAGTGATGACAAACCTGTTCTGCTCCTGATCTCGTTCGCGATGGCTACATGT	2989	TITLE			
Db	901	TTGACCTCAGCAGCAGCAACCTGGTTCTGCCCCCTGATCTCGTTCCTGGCGGTACATGT	960	JOURNAL			
Qy	2990	GGGAATGCCCAACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGACCTGC	3049	FEATURES			
Db	961	GGGAATGCCCAACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGACCTGC	1020	Source			
Qy	3050	TGATTTTCTGTCCACAAGGATTTGGACCAATCCAGCATGAGGTTACTCACTACCAAGCT	3109	CDS			
Db	1021	TGATTTTCTGTCCACAAGGATTTGGACCGTATCGATGATGAGTTACTCACTACCAAGCT	1080				
Qy	3110	CTGACAGTCGGGATGATGCTGCACAGCTTGAAGAACACCTCTCGCGTCTGGAG	3169				
Db	1081	CTGACAGTCGGGATGATGCTGCACAGCTTGAAGAACACCTCTCGCTCGCTGGAG	1140				
Qy	3170	GATTCAGCAGCTGGATTTCCGGCCATGAATCTACGCCACCGCAGGTTCCAGTAAACGGTT	3229				
Db	1141	GATTCAGCAGCTGGATTTCCGGTCATGAATCTACGCCCGCAGGTTGCAGTCAACGGTT	1200				
Qy	3230	CTGATCGCTGGCTCGTGGGCTGGATGGGGCTCGCGCGCAGAGATCAACCCACAGTTG	3289				
Db	1201	CCGATCGCTGGCTTGTGGGCTGGATGGGATTCCTCGCACAGAGATCAACCCACAGTTG	1260				
Qy	3290	CACAGAAGGATGGGTGCACCTGCCTGACTGTCGCCCGCAAGCTTCATTTGGCGAACACG	3349				
Db	1261	CGCAGAAGGATGGGTGACCTGCCTGACCGCTTCCTCGCAGGCTCATTTGGGTACCATG	1320				
Qy	3350	CGATCTACCAAGAGCTCTCTTCCACAGAGGGGATCGGGGGTAATCAGATCTGTATTAG	3409				
Db	1321	CGATCTATCAAGAGCTCTCTTCCACAGAGGGGAGTCGGGGGTAACTAGATCTGTATTAG	1380				
Qy	3410	GTTCGTAACCTGTCAGATAGACATCCAGAGCAATATTTCCCTCGAGTGGATGCTGTC	3469				
Db	1381	GTTCGTAACCTGTCAGATAGACATCCGAGCAATGTTTCCCTCGAGTGGATGGTGTCC	1440				
Qy	3470	GTTTGTCTGTGTGATGATGGTGTGATCGTCGCGTAGCTGAGGTAAACCTGCGCAATTAG	3529				
Db	1441	GGTTGTCTGTGTGATGCGGATGCGGATCGTCGTGTAGCTGAAGTAAACCTGCGCAATTAG	1500				
Qy	3530	TGATCGGGACGATATACAGCCATTTGAGATTAAGTCAGAGGTGATGACAGGTTTCATCG	3589				
Db	1501	TGATCGGGACGATATACAGCCATTTGAGATTAAGTAAACAGAGGTGATGCGCAGGTTTCC	1560				
Qy	3590	CTTTTCCGGGCGCTTCAAAGGTGACACTATTGAGAGATAGTATATAAAAGGCTCTTTTG	3649				



QY	4558	AGACTTCTGATCACCCAGTGTGACTCTGCTGTGCTCACCGCTTCCTTACGTTCTATTGC	46117
Db	301	AGACTTCCTGATCACTCCAGTGTGCTACGTTGTGCTCACCGGATTCCTTACATTCATCGC	360
QY	4618	TATTGTCCAGCAATCGCTGGTGGTCACTTGTCTGGCCACACGCTCTCAGGGACTCTA	4677
Db	361	CATTGGCCACGAATGCGTGGTGGGCGATGTCTGGCCACACGGTCTACAGGACTTTA	420
QY	4678	TGATTTGGTGGTCCAGTCCGGGGTCTGCTTTTCGCTCTGGCTTACTACCAATGTTAT	4737
Db	421	TGATTTCCGTTGGTCCAGTCCGGGGTCTGCTCTTCGCTCTGGCTTACTACCAATGCTCAT	480
QY	4738	CACGTGCTGCACACGTCCTTCCGGCCAATTGAGCTGGAGCTGTTCAACACAGGGTGGATC	4797
Db	481	CACGTGCTGCACACGTCCTTCCGGCCAATTGAGCTGGAGCTGTTTAAACAGGGTGGATC	540
QY	4798	CTTCATCTTGGCAACCGCATPCATGGCCAAATPCGGCAGGGTGCAGCATGTTTGGCAGT	4857
Db	541	CTTCATCTTGGCAACCGCATCTATGCTTAATPCGGCCAGGGTGCAGCATGTTTGGCAGT	600
QY	4858	GTTCCTCTTAGCGAAGAGTGAAGAGCTCAAGGSCCTTGACGGTGTCTCAGGTGCTCTCCGC	4917
Db	601	GTTCCTCTTAGCGAAGAGTGAAGAGCTCAAGGSCCTTGACGGTGTCTCAGGTGCTCTCCGC	660
QY	4918	TGTTCTTGGTATACAGAGCCTCGATCTTCGGTGTGAACCTTCGGCTCGGCTGGGCTGGCTT	4977
Db	661	TGTTCTTGGTATACAGAGCCTCGATCTTCGGTGTGAACCTTCGGCTCGGCTGGGCTGGCTT	720
QY	4978	CTACATTTGGTATCCGTACCCAGCATCTCGGTGGCGCTTTGATTGCACCTCTTTTCATATCAA	5037
Db	721	CTTCATCCGTTATCGGTACCCAGCATCTCGGTGGCGCTTTGATTGCACCTCTTTTAAATCAA	760
QY	5038	GGCAGTTGCTTGGGGCTGCGAGTTTCTTGGGTGTGTTTCTATTGATGCTCCAGATAT	5097
Db	781	GGCAGTTGCTTGGGGCTGCGAGTTTCTTGGGTGTGTTTCTATTGATGCTCCAGATAT	840
QY	5098	GGTCATGTTCTTGGTTTGGCGGTAGTTACCTTTGTCACTCGCATTCGGCCAGCAGATTGC	5157
Db	841	GGTCATGTTCTTGGTGTGTCAGTTGTTACCTTCTTCACTCGCATTCGGCCAGGATTGC	900
QY	5158	TTATGGCCTTTACTTGGTTTCGCGCAACGGCAGCATTTGATCCAGATGCAACCGTGTCTC	5217
Db	901	TTATGGCCTTTACTTGGTTTCGCGCAACGGCAGCATTTGATCCAGATGCAACCGTGTCTC	960
QY	5218	AGTGCTCTCAGGAACGACCAAGCGAAGCAGACCCGCGAGATTTTCAACGATTC	5277
Db	961	AGTGCTCTCAGGAACGACCAAGCGAAGCAGACCCGCGAGATTTTCAACGATTC	1020
QY	5278	CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTCACGATGC	5337
Db	1021	CACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTCGCACTGAGCAGCGTCACGATGC	1080
QY	5338	CATGTTTCCAGCGGAAAGCTTGGCTCAGGTGTTTCGATCGTCCCAACCAAGGGGAGCT	5397
Db	1081	CATGTTTCCAGCGGAAAGCTTGGCTCGGCGTTGCCATCGTCCCAACCAAGGGGAGCT	1140
QY	5398	GGTTTCACCAAGTGCAGGAAGATCTGGTGGCTTCCCATCTGGTTCACCGTTTCGAGT	5457
Db	1141	AGTTTCTCCGGTGAAGTGAAGATTTGGTGGCATTCCTCCATCTGGCCATGCTTTTCGAGT	1200
QY	5458	CCGCACTTAAGCTGAGGATGGTTCCAAATGGATATCTTGATGCACATTTGGTTTCGACAC	5517
Db	1201	TCGCAACCAAGCTGAGGATGGTTCCAAATGGATATCTTGATGCACATTTGGTTTCGACAC	1260
QY	5518	CGTAAACCTCAACGGCAGCATTACCCCGCTGAAGAACGAGGGCATGAAGTCAAGC	5577
Db	1261	AGTAAACCTCAACGGCAGCATTACCCCGCTGAAGAACGAGGGCATGAAGTCAAGC	1320
QY	5578	AGGGAGCTGCTGTGTGAATTCGATATTTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC	5637
Db	1321	AGGGAGCTGCTGTGTGAATTCGATATTTGATGCCATTAAGGCTGCAGGTTATGAGGTAAC	1380

Qy	5638	CAGCCGATGTGGTTTCGAATTACAAAGAAAACCAGCACTGTAAACACITACCGTITGGG	5691
Db	1381	CAGCCGATGTGGTTTCGAATTACAAAGAAAACCAGCACTGTAAACACITACCGTITGGG	1440
Qy	5698	CGAAATTAAGCGGGAGCACAACCTTGCTCAACGTCGCAAAAGAAAAGAGCGGTGCCAGCAC	5757
Db	1441	CGAATTTGAAGCGGGAGCACAACCTTGCTCAACGTCGCAAAAGAAAAGAGCGGTGCCAGCAC	1500
Qy	5758	ACCATAGTTGAACACTTGAGTGTTCG	5784
Db	1501	ACCATAGTTGAACACTTGAGTGTTCG	1527
RESULT 11			
AX065221		1342 bp DNA linear PAT 24-JAN-2001	
LOCUS	AX065221	Sequence 347 from Patent WO0100844.	
DEFINITION	AX065221	AX065221.1 GI:12542933	
ACCESSION	AX065221		
VERSION	AX065221		
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
.source			
		1..1342	
	/organism="Corynebacterium glutamicum"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:1718"		
	101..>1342		
	/note="RXA00041"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="CAC25850.1"		
	/db_xref="GI:12542934"		
	/translation="MHTLSRLPAYHVTPEGLNDPNGMYVDGDTLHVYXOHPDGF		
	FPAPKRGWAHTTPTLPGLQRQLTHPLDALYDLDCGSCSGVAFPTDRDPMSIDF		
	TGNLKIDGRKAHTNLEVEDPTGLMGIIHRSPKNEIIDGPASGFTPHTRDMPSIDF		
	GCKMKVLVRGEENTGALAAVLYRSTDLLENFEISGETTFDSDDQGPGAPDLVPBGYMMW		
	ECPNLFILTOEETGELDVLIIFCPQGLDIRHDEVIHYASSDQQGVYVKLEGTFRVLL		
	RFSSELDFFHEFTPAQVAVGSDAWLVGMWGMLPAQDDHPTVAREGWYHCILTVPKRKLHL		
	RNHAIYOELLPESEGIVRSVLGSEPVRVIDRGNISLEWDGVRLSVDRGDRAAEV		
	KPGELVTADDNTATEITAGDQ"		
BASE COUNT	312 a 373 c 359 g 298 t		
ORIGIN			
Query Match	22.1%; Score 1321.2; DB 6; Length 1342;		
Best Local Similarity	99.0%; Pred. No. 0;		
Matches 1329; Conservative	0; Mismatches 13; Indels 0; Gaps 0;		
Qy	2238	ATGAAGCAGCAGTAGTCCAAGCTGGAAAAACGCTGATCACTACCGTCTCATGGAGCAATAA	2297
Db	1	ATGAAGCAGCAGCATCCAAGCTGGAAACGCTGATCACTACCGTCTCATGGAGCAATAA	60
Qy	2298	AGCTGCCTAGAAACAAAGAAAGTACTGTGTGGGCTATGCACACAGAACCTTCCAG	2357
Db	61	AGCTGCCTAGAAACAAAGAAAGTACTGTGTGGGCTATGCACACAGAACCTTCCAG	120
Qy	2358	TTTGGCCCTCGGTACCATGTGACTCCTCCGAGGCGAGGCTCAATGATCCCAACGGAAT	2417
Db	121	TTTGGCCCTCGGTACCATGTGACTCCTCCGAGGCGAGGCTCAATGATCCCAACGGAAT	180
Qy	2418	GTACGTGATGGAGATACCCCTCCACGTCTACTACAGCAGCATCCAGGTTCCGCTTCGC	2477
Db	181	GTACGTGATGGAGATACCCCTCCACGTCTACTACAGCAGCATCCAGGTTCCGCTTCGC	240
Qy	2478	ACCAAAGCGACCGGCTCGGGCTCAACACCACGACCGCTTGTACCGGACCGCAGGATTGCA	2531

```
|||||
241 ACCAAGCGCAGCGCTGCGCTCACACACACGCGGTTGACGGACCGCAGGATTGCA 300
2538 GTGAGCGACCTGCGGAGCTCTTACCGGATGATCCTATGACCTGGATGGATGCTA 2597
|||||
301 GTGAGCGACCTGCGGAGCTCTTACCGGATGATCCTATGACCTGGATGGATGCTA 360
2598 TTCGGTGGAGCGGTATTTACTGACGGACACTTAAACTTTTCTACACGGCAACCTAAA 2657
|||||
361 TTCGGTGGAGCGGTATTTACTGACGGACACTTAAACTTTTCTACACGGCAACCTAAA 420
2658 AATTGACGGAAGCGCGCCACCAACCTTGCAGAGTCGAGGACCCCACTGGGT 2717
|||||
421 AATTGACGGAAGCGCGCCACCAACCTTGCAGAGTCGAGGACCCCACTGGGT 480
2718 GATGGCGGCAATCATCGCGTTCGCTTAAATTCGCTTATCGAGCGACCCGCGCGG 2777
|||||
481 GATGGCGGCAATCATCGCGTTCGCTTAAATTCGCTTATCGAGCGACCCGCGCGG 540
2778 TTTACACCCCATACCGGATCCCATGATCAGCCTGATGTTGATGTTGAAACATGTT 2837
|||||
541 TTTACACCCCATACCGGATCCCATGATCAGCCTGATGTTGATGTTGAAACATGTT 600
2838 TCTTGGGCGCAACGCGCAACCTACCGGTGAGCGGTTCTATACCGCTGCACAGATCT 2897
|||||
601 TCTTGGGCGCAACGCGCAACCTACCGGTGAGCGGTTCTATACCGCTGCACAGATCT 660
2898 TGAACCTGGGAATCTCCGGTGAATACCTTTGACCTCAGTGATGACACACCTGGTTC 2957
|||||
661 TGAACCTGGGAATCTCCGGTGAATACCTTTGACCTCAGTGATGACACACCTGGTTC 720
2958 TGTCTCTGATCTCGTCCCGATGGCTACATGCGGAATGCCCAACCTTTTACGGTTCG 3017
|||||
721 TGTCTCTGATCTCGTCCCGATGGCTACATGCGGAATGCCCAACCTTTTACGGTTCG 780
3018 CGATGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCACAAAGGATGGACCG 3077
|||||
781 CGATGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCACAAAGGATGGACCG 840
3078 AATCCAGCATGAGTTACTCATTACGCAAGCTCTGACGAGTGGGATGTTCTGTCACAA 3137
|||||
841 AATCCAGCATGAGTTACTCATTACGCAAGCTCTGACGAGTGGGATGTTCTGTCACAA 900
3138 GCTTGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCACAAAGGATGGACCG 3197
|||||
901 GCTTGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCACAAAGGATGGACCG 960
3198 ATTCTACGACCGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3257
|||||
961 ATTCTACGACCGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
3258 GCTGCCCGCAGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3317
|||||
1021 GCTGCCCGCAGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
3318 TGTGCCCGCAGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3377
|||||
1081 TGTGCCCGCAGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
3378 GGGGAGTCCGGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3437
|||||
1141 GGGGAGTCCGGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
3438 AGGCAATATTTCCCTCAGTGGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3497
|||||
1201 AGGCAATATTTCCCTCAGTGGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
3498 TCGGAGTCCGGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3557
|||||
1261 TCGGAGTCCGGGATGATCAACGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
3558 GATTAAGTCCGGGATGATGACAG 3579
|||||
```

Db 1321 GATAACTGCAGGTGATGGACAG 1342

RESULT 12  
AX122989 1299 bp DNA linear PAT 11-MAY-2001

LOCUS AX122989  
DEFINITION Sequence 2905 from Patent EP1108790.  
ACCESSION AX122989  
VERSION AX122989.1 GI:14040477  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.

1  
Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
Novel polynucleotides  
Patent: EP 1108790-A 2905 20-JUN-2001;  
KIOWA HAKKO KOGYO CO., LTD. (JP)  
Location/Qualifiers  
1..1299  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"

BASE COUNT 287 a 363 c 350 g 299 t  
ORIGIN

Query Match 21.2%; Score 1267.8; DB 6; Length 1299;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1286; Conservative 0; Mismatches 12; Indels 1; Gaps 12;

QY 2327 TGTGTGGGGCTATGCACACAGAACTTCCAGTTTGGGCCCTGGTACCATGTGACTCTC 2386  
Db 2 TGTGTGGGGCTATGCACACAGAACTTTCAGTTTGGGCCCTGGTACCATGTGACTCTC 61  
QY 2387 CGCAGGGAGGCTCAATGATCCCAACGGAATGTAGTCGATGGAGATACCTCCACGCT 2446  
Db 62 CGCAGGGAGGCTCAATGATCCCAACGGAATGTAGTCGATGGAGATACCTCCACGCT 121  
QY 2447 ACTACACAGATCCAGTTTCCCTTCGCACCAAGCGCACCGCTGGGCTCACACCA 2506  
Db 122 ACTACACAGATCCAGTTTCCCTTCGCACCAAGCGCACCGCTGGGCTCACACCA 181  
QY 2507 CCACGCCGTTGACCGGACCGCAGCGATTGCACTGCGAGCAGCTGCGCGACGCTCTTACC 2566  
Db 182 CCACGCCGTTGACCGGACCGCAGCGATTGCACTGCGAGCAGCTGCGCGACGCTCTTACC 241  
QY 2567 CGGATGCATCTATGACCTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2626  
Db 242 CGGATGCATCTATGACCTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 301  
QY 2627 CACTTAAACTTTTCTACACCGGACCTTAAATTTGACGGAAGCGCGCCACCCAAA 2686  
Db 302 CACTTAAACTTTTCTACACCGGACCTTAAATTTGACGGAAGCGCGCCACCCAAA 361  
QY 2687 ACCTTGTGAAGTCCAGACCCAACTGGGCTGATGGCGGATTCATGCGGCTTGCCTA 2746  
Db 362 ACCTTGTGAAGTCCAGACCCAACTGGGCTGATGGCGGATTCATGCGGCTTGCCTA 421  
QY 2747 AAAATCCGCTTATGACGAGCCCGCAGCGGTTTACACCCCATATACCGGATCCATGA 2806  
Db 422 AAAATCCGCTTATGACGAGCCCGCAGCGGTTTACACCCCATATACCGGATCCATGA 481  
QY 2807 TCAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2866  
Db 482 TCAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 541  
QY 2867 GTGACGCGGTTCTATACCGCTCGACAGATCTTGAACCTGGGAATCTCCGGTGAATCA 2926  
Db 542 GTGACGCGGTTCTATACCGCTCGACAGATCTTGAACCTGGGAATCTCCGGTGAATCA 601  
QY 2927 CCTTTGACCTCAGTGTGATGACACCACTGCTTCTGCTCTGCTCTGCTCTGCTCTGCTCTA 2986

```

Db 602 CCTTTGACCTCAGTGATGACAACTGGTTCTGCTCCTGATCTCGTTCGGTGGGTACA 661
QY 2987 TGTGGGAATGCCCAACCTTTTACGCTTCGCATGAAGAACTGGCGAAGATCTCGACG 3046
Db 562 TGTGGGAATGCCCAACCTTTTACGCTTCGCATGAAGAACTGGCGAAGATCTCGACG 721
QY 3047 TGTGATTTCTGTCCACAGGATTTGACCGAATCCACGATGACGATGAGTTACTACTAGCAA 3106
Db 722 TGTGATTTCTGTCCACAGGATTTGACCGAATCCACGATGACGATGAGTTACTACTAGCAA 781
QY 3107 GCTCTGACAGTGGGATATGTGTCGACAAAGTTGAAGAACTGACCTCCGGGTCTTGC 3166
Db 782 GCTCTGACAGTGGGATATGTGTCGACAAAGTTGAAGAACTGACCTCCGGGTCTTGC 841
QY 3167 GAGGATTCAGGAGCTGGATTTGGGCCATGAATCTACGACCGCAGGTTGCAGTAAAG 3226
Db 842 GAGGATTCAGGAGCTGGATTTGGGCCATGAATCTACGACCGCAGGTTGCAGTAAAG 901
QY 3227 GTTCTGATGCTGGCTGCTGGGTGGATGGGCTGCCCGCAGGATGATCACCAACAG 3286
Db 902 GTTCTGATGCTGGCTGCTGGGTGGATGGGCTGCCCGCAGGATGATCACCAACAG 961
QY 3287 TTGCACAGGAAGATGGTGCACCTGCTGACTGTGCCCGCAAGCTTCATTTCCGCAACC 3346
Db 962 TTGCACAGGAAGATGGTGCACCTGCTGACTGTGCCCGCAAGCTTCATTTCCGCAACC 1021
QY 3347 ACGGGATCTACCAAGACTCCTTCTCCACAGGGGAGTCGGGGTAATCAGATCTGTAT 3406
Db 1022 ACGGGATCTACCAAGACTCCTTCTCCACAGGGGAGTCGGGGTAATCAGATCTGTAT 1081
QY 3407 TAGTCTTGAACCTGTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTG 3466
Db 1082 TAGTCTTGAACCTGTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTG 1141
QY 3467 TCCGTTTGTCTGTGATCGTGTGATGATCGTCGCTAGCTGAGGTAAACCTTGGCGAAT 3526
Db 1142 TCCGTTTGTCTGTGATCGTGTGATGATCGTCGCTAGCTGAGGTAAACCTTGGCGAAT 1201
QY 3527 TAGTGATCGGGAGGATTAACAGCCATTGAGATAACTGCAGGTGATGGACAGTTTCAT 3586
Db 1202 TAGTGATCGGGAGGATTAACAGCCATTGAGATAACTGCAGGTGATGGACAGTTTCAT 1261
QY 3587 TCGCTTTTCCGGGCTTCAAGGTGACACTATTGAGAGA 3625
Db 1262 TCGC-TTTCGGGGCTTCAAGGTGACACTATTGAGAGA 1299

RESULT 13
BD165106
LOCUS BD165106 1299 bp DNA linear PAT 17-JAN-2003
DEFINITION Novel polynucleotide.
ACCESSION BD165106
VERSION BD165106.1 GI:27870918
KEYWORDS JP 2002191370-A/2905.
SOURCE unidentified
ORGANISM Corynebacterium glutamicum
REFERENCE
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoo,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2905 09-JUL-2002;
KYOWA HAKKO KOGYO CO LTD
COMMENT OS Corynebacterium glutamicum
PN JP 2002191370-A/2905
PD 09-JUL-2002
PF 15-DEC-2000 JP 2000405096
PI SATOSHI NAKAGAWA,HIROSHI MIZOGUCHI,SEIKO ANDO,MIKIO HAYASHI,
PI KEIKO OCHIAI,
PI HARUHIKO YOKOI,NAOKO TATEISHI,AKIHIRO SENOO,MASATO IKEDA,AKIO
PI OZAKI
PC C12N15/09,C12N15/09,C07K14/34,C07K16/12,C07K16/40,C12M1/00, PC
```

```

C12N1/15,
PC C12N1/19,C12N1/21,C12N5/10,C12N9/00,C12N9/02,C12P7/40,C12P13/
PC 04,C12P13/08,
PC C12P19/00,C12P19/34,C12P21/02,C12Q1/37,C12Q1/68,G01N33/53, PC
G01N33/566,
PC G01N33/569,G01N33/68,G01N37/00//C12P21/08,(C12N1/21,C12R1/15),
PC (C12N1/21,C12R1/13),(C12N1/21,C12R1/01),(C12P13/08,C12R1/15),
PC C12N15/00,
PC C12N5/00,C12N15/00
CC Novel polynucleotide
FH Key Location/Qualifiers
FT source 1..1299 /organism='Corynebacterium glutamicum'
FT Location/Qualifiers
1..1299
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 287 a 363 c 350 g 299 t
ORIGIN
Query Match 21.2%; Score 1267.8; DB 6; Length 1299;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
QY 2327 TGTGTGGGCTATGCACACAGAACTTCCAGTTTGGCCCTCGTACCATGTGACTCCTC 2386
Db 2 TGTGTGGGCTATGCACACAGAACTTCCAGTTTGGCCCTCGTACCATGTGACTCCTC 61
QY 2387 CGCAGGCGAGGCTCAATGATCCCAACGGAATAGCTGATGGCGCATCCCTCCAGTCT 2446
Db 62 CGCAGGCGAGGCTCAATGATCCCAACGGAATAGCTGATGGCGCATCCCTCCAGTCT 121
QY 2447 ACTACGACGACGATCCAGGTTTCCCTTCGCACCAAGCGACCGGCTGGGCTCACACCA 2506
Db 122 ACTACGACGACGATCCAGGTTTCCCTTCGCACCAAGCGACCGGCTGGGCTCACACCA 181
QY 2507 CCACGCGGTTGACCGGACCGGACGATTCAGTGGACGACCTGCCGACGCTCTTTACC 2566
Db 182 CCACGCGGTTGACCGGACCGGACGATTCAGTGGACGACCTGCCGACGCTCTTTACC 241
QY 2567 CGGATGCATCTTATGACCTGGATGGATGCTATTCGGTGGAGCCGCTATTACTAGCGCA 2626
Db 242 CGGATGCATCTTATGACCTGGATGGATGCTATTCGGTGGAGCCGCTATTACTAGCGCA 301
QY 2627 CACTTAACTTTTACACCGGCAACCTTAAATTTGACGGAAGCGCGGCCACCCCAA 2686
Db 302 CACTTAACTTTTACACCGGCAACCTTAAATTTGACGGAAGCGCGGCCACCCCAA 361
QY 2687 ACCTTGTGAAGTCGAGGACCCCACTGGGCTGATGGGCGCATTCATCGCGTTTCGCTA 2746
Db 362 ACCTTGTGAAGTCGAGGACCCCACTGGGCTGATGGGCGCATTCATCGCGTTTCGCTA 421
QY 2747 AAAATCCGCTTATGACGAGCCCGGAGGTTTTCACACCCCATTTACCGGATCCCATGA 2806
Db 422 AAAATCCGCTTATGACGAGCCCGGAGGTTTTCACACCCCATTTACCGGATCCCATGA 481
QY 2807 TCAGCGCTGATGATGATGTTTGAACATGTTTGGGGCCCAACGCGAAACCTTCACCG 2866
Db 482 TCAGCGCTGATGATGATGTTTGAACATGTTTGGGGCCCAACGCGAAACCTTCACCG 541
QY 2867 GTGCAGCGGTTCTATACCGCTCGACAGATCTTGAACCTGGGAATTCCTCGGTGAATCA 2926
Db 542 GTGCAGCGGTTCTATACCGCTCGACAGATCTTGAACCTGGGAATTCCTCGGTGAATCA 601
QY 2927 CTTTTCACCTCAGTGTGACACACCTGGTTCTGCTCTGATCTCCCTCCGATGCTACA 2986
Db 602 CTTTTCACCTCAGTGTGACACACCTGGTTCTGCTCTGATCTCCCTCCGATGCTACA 661
QY 2987 TGTGGGAATGCCCAACCTTTTACGCTTCGCGATGAGAAACTGCGGAAGATCTCGACG 3046
Db 662 TGTGGGAATGCCCAACCTTTTACGCTTCGCGATGAGAAACTGCGGAAGATCTCGACG 721
```

```
QY 3047 TGTGATTTTCTGTCACAAAGGATGGACCGAATCCAGATGAGGTACTACTACGCAA 3106
Db 722 TGTGATTTTCTGTCACAAAGGATGGACCGAATCCAGATGAGGTACTACTACGCAA 781
QY 3107 GCTCTGACCAAGTGGGATATGTCGCGACAAGCTTGAAGGAAGACCTTCGCGTCTTGC 3166
Db 782 GCTCTGACCAAGTGGGATATGTCGCGACAAGCTTGAAGGAAGACCTTCGCGTCTTGC 841
QY 3167 GAGGATTCAGCAGCTGGATTTTCGGCCATGAATTTCTACGACCGCAGGTTCAGTAACG 3226
Db 842 GAGGATTCAGCAGCTGGATTTTCGGCCATGAATTTCTACGACCGCAGGTTCAGTAACG 901
QY 3227 GTTCTGATGCTGGCTGCTGGGCTGGATGGGCTGCGCGCAGGATGATACCCAAACG 3286
Db 902 GTTCTGATGCTGGCTGCTGGGCTGGATGGGCTGCGCGCAGGATGATACCCAAACG 961
QY 3287 TTGCACAGAAAGGATGGGTGCACTGCTGACATGTCGCCGCAAGTTTCAATTTGCCAAC 3346
Db 962 TTGCACAGAAAGGATGGGTGCACTGCTGACATGTCGCCGCAAGTTTCAATTTGCCAAC 1021
QY 3347 ACAGCATACCAAGAGCTCTTCTCCAGAGGGGAGTCGGGGTAAATCAGATCTGTAT 3406
Db 1022 ACAGCATATCAAGAGCTCTTCTCCAGAGGGGAGTCAGGGTAAATCAGATCTGTAT 1081
QY 3407 TAGGTTCTGAAGCTGTCGAGTAGACATCGGAGGCAATATTTCCCTCGAGTGGGATGG 3466
Db 1082 TAGGTTCTGAAGCTGTCGAGTAGACATCGGAGGCAATATTTCCCTCGAGTGGGATGG 1141
QY 3467 TCGTTTGTCTGTGGATCGTGATGATGTCGCGGTAGCTGAGGTAAACCTGGCGAAT 3526
Db 1142 TCGTTTGTCTGTGGATCGTGATGATGTCGCGGTAGCTGAGGTAAACCTGGCGAAT 1201
QY 3527 TAGTGATCGGACGATATACAGCATTGAGTACTGAGTACTGAGTACTGAGTACTGAGT 3586
Db 1202 TAGTGATCGGACGATATACAGCATTGAGTACTGAGTACTGAGTACTGAGTACTGAGT 1261
QY 3587 TCGTTTTCGGGCTTCAAAGTGACACTATTGAGAGA 3625
Db 1262 TCGC-TTTCGGGCTTCAAAGTGACACTATTGAGAGA 1299

RESULT 14
AX065287 1287 bp DNA linear PAT 24-JAN-2001
LOCUS AX065287
DEFINITION Sequence 413 from Patent WO0100844.
ACCESSION AX065287
VERSION AX065287.1 GI:12542999
KEYWORDS Corynebacterium glutamicum
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
carbon metabolism and energy production
JOURNAL Patent: WO 0100844-A 413 04-JAN-2001;
BASIS BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
    Location/Qualifiers
        1..1287
            /organism="Corynebacterium glutamicum"
            /mol_type="genomic DNA"
            /db_xref="taxon:1718"
            101..1267
                /note="RXN00043"
                /codon_start=1
                /transl_table=11
                /db_xref="GI:12543000"
                /translation="MAEVHYOENAGQAVKIEGRIVTPHGVDFGLQLENGIITELS
                GERAPXNAGFHPLEPFIVSFIDLHNEGGNGAFPTGTDQDQARNQAQVHREGTVMFL
                ASKVPADALAPLQELGIEGLGIEGPFINACRGANQPFIFPQNTDL
                AQVIRHAGKWNKISITVAPETDNLTELLD:CAAHBHIIASFQHTDADFDFTTSAIAAKE
                CDS
```

```
BASE COUNT 325 a 349 c 326 g 287 t
ORIGIN
Query Match 20.7%; Score 1237.4; DB 6; Length 1287;
Best Local Similarity 97.6%; Pred. No. 1.4e-306;
Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 242 AACAGCAGGCTCAAGTCGGAAGATAATTAACCTAGATCCGTAGACATAAGACATCATAC 301
Db 1 AACAGCAGGCTCAAGTCGGAAGATAATTAACCTAAATCCGTAGACATAAGACATCATAC 60
QY 302 GTCCTATGCTTGTCTGGAAGAACCAATAACCTCAGAAAGATGGCAGAAAGTGTGCAATTA 361
Db 61 GTCCTATGCTTGTCTGGAAGAACCAATAACCTCAGAAAGATGGCAGAAAGTGTGCAATTA 120
QY 362 TCAGAAAAATCAGGTCAAGCAGTTAAAAAATTTAGGGAAGATTTTCCCCCTCCG 421
Db 121 TCAGAAAAATCAGGTCAAGCAGTTAAAAAATTTAGGGAAGATTTTACCCCCCAGG 180
QY 422 GGTGATTGATGGCTTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGAGAAACC 481
Db 181 GGTGATTGATGGCTTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGAGAAACC 240
QY 482 AGCACCCTAAAAACGAGGATTTCCACCCGAACTCCCAAGATTTTCCGGTTTATTGA 541
Db 241 AGCACCCTAAAAACGAGGATTTCCACCCGAACTCCCAAGATTTTCCCGAGTTTATTGA 300
QY 542 TCTTCATTAATCAGGTGGAACCGTGGCGCTTTCTACGGGAACGACGACGACGAGCGAG 601
Db 301 TCTTCATTAATCAGGTGGAACCGTGGCGCTTTCTACGGGAACGACGACGACGAGCGAG 360
QY 602 GAACACCGCGCAGTATCACCCGGAACATGGCAGACCGTGTATGTTGCCAGCATGTTTC 661
Db 361 GAATGCCCGCAGTATCACCCGGAACATGGCAGACCGTGTATGTTGCCAGCATGTTTC 420
QY 662 GCGCGCGCTGACGCACCTGGCAGCGCAGGTGGAAGAACTTATTCCTTGTGTAAGAGGT 721
Db 421 GCGCGCGCTGACGCACCTGGCAGCGCAGGTGGAAGAACTTATTCCTTGTGTAAGAGGT 480
QY 722 CCTGCTGTGCGGCATTCACCTCGAGGGGCTTTTCATCAACGCATGCGTGTGTTGCTCA 781
Db 481 CCTGCTGTGCGGCATTCACCTCGAGGGTCTTTTCATCAACGCATGCGTGTGTTGCTCA 540
QY 782 AAACCCGGATTTTATTTTCCGCAACCCCAACAGATCTTGCCAGGTGATCCATGCGGG 841
Db 541 AAACCCGGATTTTATTTTCCGCAACCCCAACAGATCTTGCCAGGTGATCCATGCGGG 600
QY 842 AAAAGGTTGATCAAAATCGATCAGTAGCGCGGAACTGACAACTTTCTTGAGCTTCT 901
Db 601 AAAAGGTTGATCAAAATCGATCAGTAGCGCGGAACTGACAACTTTCTTGAGCTTCT 660
QY 902 CGATCTCTGCCAGCGCACCATCATCTTCTTCCTCGGGCAGCTGATGACAGATTTTGA 961
Db 661 CGATCTCTGCCAGCGCACCATCATCTTCTTCCTCGGGCAGCTGATGACAGATTTTGA 720
QY 962 TACCACCTACCAGCGCAATTCCTTGGCTTAAAGAGAAAAATGTGACGCTACGGGTACGCA 1021
Db 721 TACCACCTACCAGCGCAATTCCTTGGCTTAAAGAGAAAAATGTGACGCTACGGGTACGCA 780
QY 1022 TTTGTTCAATGGATGCTCGCTGCATCATAGGCTCCCGGACGCTGGCGCTTTGCT 1081
Db 781 TTTGTTCAATGGATGCTCGCTGCATCATAGGCTCCCGGACGCTGGCGCTTTGCT 840
QY 1082 TGCTGCGGCACGTCCGCGGGACGCATATGTTAGTTGATGCCGACGCGCTGTCATTTGGC 1141
Db 841 TGCTGCGGCACGTCCGCGGGACGCATATGTTAGTTGATGCCGACGCGCTGTCATTTGGC 900
QY 1142 CGATGGAACGTGTCATCTAGCTCGTTCACAAACGCCCTTTTTCATCAGGACGCGCATGGA 1201
Db 901 CGATGGAACGTGTCATCTAGCTCGTTCACAAACGCCCTTTTTCATCAGGACGCGCATGGA 960
```



Db 1141 AACCGTCGGCCAAATTCGAGCTTAGCGATCACGAAATCGTTAAATCCAAACCCTGT 1200  
Qy 1442 AAATTTTGTGGTCTTTGACTCAAACGGCCAGGTGCAAAAGGTCCATTAGGTCAATCAAGT 1501  
Db 1201 AAATTTTGTGGTCTTTGACTCAAACGGCCAGTTACAACAGGTCCATTAGACCATCAAGT 1260  
Qy 1502 ACTTTAAGTACGAGTAAACTATCCTG 1528  
Db 1261 AATTAAATACGAGCAAAACTTTCCTG 1287

Search completed: September 30, 2003, 23:16:27  
Job time : 14075 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 15:17:02 ; Search time 928 Seconds

(without alignments)  
17363.088 Million cell updates/sec

Title: US-10-019-284A-1

Perfect score: 5969

Sequence: 1 agtcctgcagccaccatt.....tgatatattgcgcgcgtgaa 5969

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c	1	5969	100.0	5969	22	AAF32543
	2	5656.2	94.8	349980	22	AAF32543
	3	3687	61.8	6911	17	AAT34177
	4	1874.2	31.4	1983	22	AAH67869
	5	1457	24.4	1656	22	AAE87497
	6	1412.6	23.7	1527	22	AAE31528
	7	1412.6	23.7	1527	22	ABG5346
	8	1321.2	22.1	1342	22	AAE71533
						Brevibacterium laci
						C glutamicum codin
						Coryneform sucrose
						C glutamicum codin
						Corynebacterium th
						C glutamicum phosph
						DNA encoding C. gl
						Corynebacterium gl

9	1267.8	21.2	1299	22	AAH67870	C glutamicum codin	
10	1237.4	20.7	1287	22	AAE71566	Corynebacterium gl	
11	1237.4	20.7	1287	22	AAE71567	Corynebacterium gl	
12	1110.4	18.6	1152	22	AAH67872	C glutamicum codin	
13	1049.8	17.6	1109	22	AAE31529	C glutamicum phosp	
14	762	12.8	882	22	AAE71571	Corynebacterium gl	
15	671.8	11.3	759	22	AAH67871	C glutamicum codin	
16	474.2	7.9	479	22	AAE7433	Corynebacterium th	
17	466.6	7.8	498	22	AAE7432	Corynebacterium th	
C	18	232.2	3.9	2913	22	AAH54100	S. epidermidis gen
C	19	215.2	3.6	1059	22	AAE71769	Corynebacterium gl
C	20	215.2	3.6	1059	23	AAE96088	C. glutamicum gene
	21	192	3.2	975	24	ABK73538	Bacillus lichenifo
	22	177.6	3.0	30246	18	AAE74367	Staphylococcus aur
	23	174.2	2.9	1380	24	ABK73394	Bacillus lichenifo
	24	173.8	2.9	2127	25	ABZ42306	Streptococcus pneu
C	25	173.8	2.9	2162598	25	ABE56454	Streptococcus pneu
	26	172.2	2.9	1965	25	ABX07668	S. pneumoniae type
C	27	172.2	2.9	3895	19	AAE52334	Streptococcus pneu
	28	163.6	2.7	1884	23	AAE55732	Streptococcus pneu
	29	161	2.7	1380	22	AAH67868	C glutamicum codin
C	30	157	2.6	11330	23	AAE55944	Propionibacterium
C	31	154.2	2.6	9769	19	AAE52163	Streptococcus pneu
	32	153	2.6	5840	20	AAE12968	Enterococcus faeca
	33	153	2.6	5840	24	ABE98763	Enterococcus faeca
	34	139.2	2.3	3037	24	ABQ70786	Enterococcus faeca
	35	139	2.3	732	24	ABE91814	Listeria monocytog
	36	137.8	2.3	1917	24	ABE67846	Staphylococcus epi
C	37	137.8	2.3	2155561	24	ABN71527	Streptococcus poly
	38	137.6	2.3	2944528	24	ABQ3041	Listeria monocytog
C	39	133.6	2.2	684707	24	ABQ67196	Listeria innocua c
	40	133.6	2.2	3011208	24	ABQ69245	Listeria innocua D
	41	122.2	2.0	567	20	AAE13601	Enterococcus faeca
	42	122.2	2.0	567	24	ABE99396	Enterococcus faeca
C	43	118.4	2.0	936	22	AAH67873	C glutamicum codin
	44	117.4	2.0	1185	25	ABZ42298	Streptococcus pneu
	45	116	1.9	1098	24	ABK73443	Bacillus lichenifo

## ALIGNMENTS

## RESULT 1

AAF32543

ID AAF32543 standard; DNA; 5969 BP.

XX AAF32543;

AC AAF32543;

XX AAF32543;

DT 20-APR-2001 (first entry)

DE Brevibacterium lactofermentum sucrose PTS enzyme II NO:1.

XX Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;

XX Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;

KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;

KW corynebacterium; phosphoenolpyruvate-sugar transport system; ds.

XX Brevibacterium lactofermentum.

OS Brevibacterium lactofermentum.

XX Brevibacterium lactofermentum.

XX Brevibacterium lactofermentum.

FT Key Location/Qualifiers

FT CDS 3779..5764

FT /\*tag= a

FT /product= "sucrose PTS enzyme II"

XX WO200102584-A1.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-JP04348.

XX 02-JUL-1999; 99JP-0189512.

XX (AJIN ) AJINOMOTO CO INC.

XX

PI · Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;  
XX  
XX WPI: 2001-138150/14.  
DR P-PSDB; AAB69080.  
XX  
XX  
PT Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme  
PT II obtained by cassette ligation-mediated amplification of downstream  
PT domain of coryneform bacterium sucrose gene, with sucrose-binding  
PT activity\_-  
XX  
XX Claim 3; Page 22-29; 45pp; Japanese.  
XX  
CC The present sequence encodes the *Brevibacterium lactofermentum* sucrose  
CC PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system or  
CC phosphoenolpyruvate-sugar transport system) enzyme II, which has  
CC sucrose-binding activity. A coryneform bacteria produced with the  
CC sucrose PTS enzyme II gene can have more efficient sugar uptake, and  
CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene  
CC and it's disrupted gene, such as one without the sucrose PTS function,  
CC can be used to produce new breeds of coryneform bacterial strains to  
CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can  
CC have improved amino-acid and nucleic acid productivity.  
XX  
SQ Sequence 5969 BP; 1388 A; 1575 C; 1578 G; 1428 T; 0 other;  
Query Match 100.0%; Score 5969; DB 22; Length 5969;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGTCGTCGAGCGCCACCATGATGTGTGTCTACCGAGCTTGGGAGGCTTCTACATCT 60  
Db 1 AGTCGTCGAGCGCCACCATGATGTGTGTCTACCGAGCTTGGGAGGCTTCTACATCT 60  
QY 61 AGCGTCCCGTCGGCGTGGAGTGGGGTCATTACGGGTGGGATCACGCCGTGAAAGTTGGC 120  
Db 61 AGCGTCCCGTCGGCGTGGAGTGGGGTCATTACGGGTGGGATCACGCCGTGAAAGTTGGC 120  
QY 121 GAACCCATGGTGTTCCTTGTGGGTGTAGGGAACAGTGCGGGTGAGAAAGTTTTCAAGTG 180  
Db 121 GAACCCATGGTGTTCCTTGTGGGTGTAGGGAACAGTGCGGGTGAGAAAGTTTTCAAGTG 180  
QY 181 TCTGCAGTTTTTAAGTTATGCATCATCAGCTTGGGAAGCTGAGGTAAATTCAGTAGACCTG 240  
Db 181 TCTGCAGTTTTTAAGTTATGCATCATCAGCTTGGGAAGCTGAGGTAAATTCAGTAGACCTG 240  
QY 241 CAACAGCAGGCTCAAGTCCGAAAGATAATTAACCTAGATCCGTAGACATAAAGACATCAT 300  
Db 241 CAACAGCAGGCTCAAGTCCGAAAGATAATTAACCTAGATCCGTAGACATAAAGACATCAT 300  
QY 301 CGTCTATGCTGTGTGGAAGGAACCAATTAACCTCAGAAAGATGGCAGAGTGGTGCAATT 360  
Db 301 CGTCTATGCTGTGTGGAAGGAACCAATTAACCTCAGAAAGATGGCAGAGTGGTGCAATT 360  
QY 361 ATCAAGAAATATGCAGGTCAAGCAGTTAAAAAATTGAGGGAAGAAATGTTCCGCCCTCG 420  
Db 361 ATCAAGAAATATGCAGGTCAAGCAGTTAAAAAATTGAGGGAAGAAATGTTCCGCCCTCG 420  
QY 421 GGGTGATTGATGGCTTTCTCCAACTCGAAACGGCATCATCAGGAACCTCTCTGGAGAAC 480  
Db 421 GGGTGATTGATGGCTTTCTCCAACTCGAAACGGCATCATCAGGAACCTCTCTGGAGAAC 480  
QY 481 CAGCACCTTAAAAACCGAGGATTCACCCCGAATCCCCAGCATTTGTCGGGTTTTATTG 540  
Db 481 CAGCACCTTAAAAACCGAGGATTCACCCCGAATCCCCAGCATTTGTCGGGTTTTATTG 540  
QY 541 ATCTTCATAATCACGGTGGAAACGGTGGCGGTTTCTACGGGAACCGCAGGACCAAGGGA 600  
Db 541 ATCTTCATAATCACGGTGGAAACGGTGGCGGTTTCTACGGGAACCGCAGGACCAAGGGA 600  
QY 601 GGAACACCGCGAGTATCACCGGAACATGCGCAGGACCGTGATGTGGCCAAAGATGGTTT 660  
Db 601 GGAACACCGCGAGTATCACCGGAACATGCGCAGGACCGTGATGTGGCCAAAGATGGTTT 660

Qy	661	CGGGCGCGGTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTCTGTGTGAAAGG	720
Db	661	CGGGCGCGGTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTCTGTGTGAAAGG	720
Qy	721	TCGTGCTGTGCGGCATTCACCTCGAGGGCCCTTTTCATCAACGCATGCGGTGTGGTGTCTC	780
Db	721	TCGTGCTGTGCGGCATTCACCTCGAGGGCCCTTTTCATCAACGCATGCGGTGTGGTGTCTC	780
Qy	781	AAACCCGGATTCATTTTCCGGCAACCCACAGATCTTGCCCGGGTGATCATCGG	840
Db	781	AAACCCGGATTCATTTTCCGGCAACCCACAGATCTTGCCCGGGTGATCATCGG	840
Qy	841	GAAAGGTTGGATCAAAATCGATCACAGTAGCGCGGAACTGACAATCTTTCTGAGCTTC	900
Db	841	GAAAGGTTGGATCAAAATCGATCACAGTAGCGCGGAACTGACAATCTTTCTGAGCTTC	900
Qy	901	TCGATCTCTCGGACGACACACATCATTTGCTTCCCTTCGGGCACATGATGCAGATTTTG	960
Db	901	TCGATCTCTCGGACGACACACATCATTTGCTTCCCTTCGGGCACATGATGCAGATTTTG	960
Qy	961	ATACCACTACAGGGCAATTCCTTGCTTAAAGAAATAATGTGACGGTCACGGCTACGC	1020
Db	961	ATACCACTACAGGGCAATTCCTTGCTTAAAGAAATAATGTGACGGTCACGGCTACGC	1020
Qy	1021	ATTGTGTTCAATTCGATGCCTCCGCTGCATCATAGGGCTCCCGGAGGGTGGCGCTTTGC	1080
Db	1021	ATTGTGTTCAATTCGATGCCTCCGCTGCATCATAGGGCTCCCGGAGGGTGGCGCTTTGC	1080
Qy	1081	TTGCTGCGGCACGTGCGGGGACGCATATGTTGAGTTGATCGCGAGCGGTGATTTGG	1140
Db	1081	TTGCTGCGGCACGTGCGGGGACGCATATGTTGAGTTGATCGCGAGCGGTGATTTGG	1140
Qy	1141	CCGATGGAAGGGTGCATCTAGCTCTGTTCCAAACAGCCTTTTTCATCACGGACGCCATGG	1200
Db	1141	CCGATGGAAGGGTGCATCTAGCTCTGTTCCAAACAGCCTTTTTCATCACGGACGCCATGG	1200
Qy	1201	AAGCGCGGGAATGCGAGACGGTGAGTACATTTTGGGGTTTTCAGAGTTCACCGTCAAG	1260
Db	1201	AAGCGCGGGAATGCGAGACGGTGAGTACATTTTGGGGTTTTCAGAGTTCACCGTCAAG	1260
Qy	1261	ATGAGTGTGCGCGCTCTCGCATGATGCGGCGCCATTCGCGGGGGACCAAGCATAGCGA	1320
Db	1261	ATGAGTGTGCGCGCTCTCGCATGATGCGGCGCCATTCGCGGGGGACCAAGCATAGCGA	1320
Qy	1321	GTCAGTTGTCGACACGTGCGGAGGGTATGAGCTTATCGACGCGACCTCCACACCT	1380
Db	1321	GTCAGTTGTCGACACGTGCGGAGGGTATGAGCTTATCGACGCGACCTCCACACCT	1380
Qy	1381	CAACCGTCGCGCTATAAATTCCTGCTCTTGGCGATCAAGAAATCGCTAAATCCAAACCTG	1440
Db	1381	CAACCGTCGCGCTATAAATTCCTGCTCTTGGCGATCAAGAAATCGCTAAATCCAAACCTG	1440
Qy	1441	CAAAATTTGTGGTCTTTGACTCAACGGCCAGGTGCAAAAGTCCATTTAGTGCATCAAG	1500
Db	1441	CAAAATTTGTGGTCTTTGACTCAAAAGGCGAGGTGCAAAAGTCCATTTAGTGCATCAAG	1500
Qy	1501	TACTTTAAGTACGAGTAAACTATCTCTGATTTTAAAGAGTCCCAACCATGGAAATCACTA	1560
Db	1501	TACTTTAAGTACGAGTAAACTATCTCTGATTTTAAAGAGTCCCAACCATGGAAATCACTA	1560
Qy	1561	TCTCAAGAGCAGCAAGATGCGCGCAAGAGTTCAGTCCCTATCGCACCTTCGCA	1620
Db	1561	TCTCAAGAGCAGCAAGATGCGCGCAAGAGTTCAGTCCCTATCGCACCTTCGCA	1620
Qy	1621	ACAAGGTTGGAACCTTTGGGGTTGCAACAGGATCCTCACACTGAGTACCTACCAAGAGC	1680
Db	1621	ACAAGGTTGGAACCTTTGGGGTTGCAACAGGATCCTCACACTGAGTACCTACCAAGAGC	1680
Qy	1681	TCATTTCGATGTGAAGCTGGGGAAGTGTCAATTCAGAGACTGCAAGGCATCTTGTTGG	1740
Db	1681	TCATTTCGATGTGAAGCTGGGGAAGTGTCAATTCAGAGACTGCAAGGCATCTTGTTGG	1740
Qy	1741	ATGAATACGTGGGACTACCCGTCAGCATGAAAACAGCTACTTTAAACCAATTCGCAAG	1800

Db	1741	ATGAATACGTGGGACTAAACCGGTGACGATGAAACACAGTACTTTAAACCAATCGCAAG	1800	Db	2821	GATGTTGGAACATGGTTCTTTGGGCCCCAACGCGGAAAAACCTCACCGGTGCGCGGTTCTA	2880
Qy	1801	AGTTCACTGACCAACATCGACATCGTTGATGAAGAGGTCTACAGCCAGATGTCGAAC	1860	Qy	2881	TACCCGCTCGACAGATCTTTGAAAACTTGGAAATCTCCGGTGAATCACCTTTGACCTCAGT	2940
Db	1801	AGTTCACTGACCAACATCGACATCGTTGATGAAGAGGTCTACAGCCAGATGTCGAAC	1860	Db	2881	TACCCGCTCGACAGATCTTTGAAAACTTGGAAATCTCCGGTGAATCACCTTTGACCTCAGT	2940
Qy	1861	CTGATCCATACGAAGCAGCTGCAGAGATGAGCAAAAGATCGCTGCAGAAATCCGTTGAAG	1920	Qy	2941	GATGCACAACCTGGTTCTGCTCCTGATCTCGTTCCCGATGGCTACATGTGGGAATGCC	3000
Db	1861	CTGATCCATACGAAGCAGCTGCAGAGATGAGCAAAAGATCGCTGCAGAAATCCGTTGAAG	1920	Db	2941	GATGCACAACCTGGTTCTGCTCCTGATCTCGTTCCCGATGGCTACATGTGGGAATGCC	3000
Qy	1921	TTCAATATCTTGATCGTGGGGAAGCGGACATCGCTTTCAATGAACCATCATCTCTCT	1980	Qy	3001	AACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATTTCTCT	3060
Db	1921	TTCAATATCTTGATCGTGGGGAAGCGGACATCGCTTTCAATGAACCATCATCTCTCT	1980	Db	3001	AACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATTTCTCT	3060
Qy	1981	GTGAGGACTGACAAAGCTGCAGCGCTGCACCCATAAAGTGTGGAGGACAAACCTCGATT	2040	Qy	3061	CCACAAGGATTGGACCGAATCCACGATGAGTTACTACTACGGAAGCTCTGACAGTGC	3120
Db	1981	GTGAGGACTGACAAAGCTGCAGCGCTGCACCCATAAAGTGTGGAGGACAAACCTCGATT	2040	Db	3061	CCACAAGGATTGGACCGAATCCACGATGAGTTACTACTACGGAAGCTCTGACAGTGC	3120
Qy	2041	CTTCACACCATCGAAGAGTCCGACCCAGCGCTGACCCAGCGTCCACCGGTTTGGGCACTTTGTC	2100	Qy	3121	GGATATGTCGTCGACAAAGCTTGAAGGAACGACCTTCCCGGCTTCCGAGGATTCAGCGAG	3180
Db	2041	CTTCACACCATCGAAGAGTCCGACCCAGCGCTGACCCAGCGTCCACCGGTTTGGGCACTTTGTC	2100	Db	3121	GGATATGTCGTCGACAAAGCTTGAAGGAACGACCTTCCCGGCTTCCGAGGATTCAGCGAG	3180
Qy	2101	CCGCGCCAAAACATCGTTGTTGGTGAAGTGTGAAGGAAAAAGCGCCATCCGCGG	2160	Qy	3181	CTGGATTTCGGCCATGAATTTACGACGCGAGGTTGAGTAAACGGTTCTGATGCCCTGG	3240
Db	2101	CCGCGCCAAAACATCGTTGTTGGTGAAGTGTGAAGGAAAAAGCGCCATCCGCGG	2160	Db	3181	CTGGATTTCGGCCATGAATTTACGACGCGAGGTTGAGTAAACGGTTCTGATGCCCTGG	3240
Qy	2161	AACGTGGAAGGCGCAGTACTGCTTCTGCTCCAGGTTCCATCCTGTAGTGCACAACT	2220	Qy	3241	CTCGTGGCTGATGGGCTGCCCGCAGGATGATCAACCAACAGTTGCACAGGAAGA	3300
Db	2161	AACGTGGAAGGCGCAGTACTGCTTCTGCTCCAGGTTCCATCCTGTAGTGCACAACT	2220	Db	3241	CTCGTGGCTGATGGGCTGCCCGCAGGATGATCAACCAACAGTTGCACAGGAAGA	3300
Qy	2221	GCCACCATCATCGTTGATGAAGCAGCAGTATCAAGCTGGAAAAAGCTGATCACTACCG	2280	Qy	3301	TGGGTGCATGCTGACTGTGCCCGCAAGCTTCATTTGGCAACCCACCGCATCACAA	3360
Db	2221	GCCACCATCATCGTTGATGAAGCAGCAGTATCAAGCTGGAAAAAGCTGATCACTACCG	2280	Db	3301	TGGGTGCATGCTGACTGTGCCCGCAAGCTTCATTTGGCAACCCACCGCATCACAA	3360
Qy	2281	TCATCATGAGCAATTAAAGCTGGCTAGAAACAAAAAGGAAAGTACTGTGGGGCTATG	2340	Qy	3361	GAGCTCCTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGATAGTTCTGAACCT	3420
Db	2281	TCATCATGAGCAATTAAAGCTGGCTAGAAACAAAAAGGAAAGTACTGTGGGGCTATG	2340	Db	3361	GAGCTCCTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGATAGTTCTGAACCT	3420
Qy	2341	CACACAGAATCTTCCAGTTTGGCGCTCGCTACCATGTACTCCTCGCGAGGCGAGCTC	2400	Qy	3421	GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTGCTGTG	3480
Db	2341	CACACAGAATCTTCCAGTTTGGCGCTCGCTACCATGTACTCCTCGCGAGGCGAGCTC	2400	Db	3421	GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTGCTGTG	3480
Qy	2401	AATGATCCCAAGGAATGTACGTGATGAGATACCTCCAGCTCTACTACCAAGCAGAT	2460	Qy	3481	GATCGTGTGATGTCGTCGCTAGCTGAGGTAAACCTGGCAATTAGTATCGCGAC	3540
Db	2401	AATGATCCCAAGGAATGTACGTGATGAGATACCTCCAGCTCTACTACCAAGCAGAT	2460	Db	3481	GATCGTGTGATGTCGTCGCTAGCTGAGGTAAACCTGGCAATTAGTATCGCGAC	3540
Qy	2461	CCAGGTTTCCCTTCGACCAAGCGCACCGCTGGGCTCACACACCGCGCTTGACC	2520	Qy	3541	GATAATACAGCCATTGAGATAACTGAGGTGATGACAGGTTTCATTCGCTTTCCGAGC	3600
Db	2461	CCAGGTTTCCCTTCGACCAAGCGCACCGCTGGGCTCACACACCGCGCTTGACC	2520	Db	3541	GATAATACAGCCATTGAGATAACTGAGGTGATGACAGGTTTCATTCGCTTTCCGAGC	3600
Qy	2521	GGACCGGAGGANTGTCAGTGGAGCGCTGCTGCCGAGCGCTTTTACCGGATGCATCCTAT	2580	Qy	3601	CTTCAAGGTGACACTATTGAGAGATAGTCATATAAAGGTCCTTTTGTGCGCAATTTGT	3660
Db	2521	GGACCGGAGGANTGTCAGTGGAGCGCTGCTGCCGAGCGCTTTTACCGGATGCATCCTAT	2580	Db	3601	CTTCAAGGTGACACTATTGAGAGATAGTCATATAAAGGTCCTTTTGTGCGCAATTTGT	3660
Qy	2581	GACCTGATGATGCTATTCCGTTGAGCGCTATTACTACGCGCACACTTAACTTTTC	2640	Qy	3661	ACAAATACCTCGCAAAATCCCTTGATCGGACACAATAAACAGGTTTAAATGTTTAGC	3720
Db	2581	GACCTGATGATGCTATTCCGTTGAGCGCTATTACTACGCGCACACTTAACTTTTC	2640	Db	3661	ACAAATACCTCGCAAAATCCCTTGATCGGACACAATAAACAGGTTTAAATGTTTAGC	3720
Qy	2641	TACACCGGCACTTAAATTTGAGGAAAGCGCGCCACCCAAACCTTTGTCGAAGTC	2700	Qy	3721	TTTTGAACAAACATTTCATGCTGAATATTTTTTGTCTTCCCGGTTAAGGAGAAATTCAT	3780
Db	2641	TACACCGGCACTTAAATTTGAGGAAAGCGCGCCACCCAAACCTTTGTCGAAGTC	2700	Db	3721	TTTTGAACAAACATTTCATGCTGAATATTTTTTGTCTTCCCGGTTAAGGAGAAATTCAT	3780
Qy	2701	GAGGACCACTGGCTGATGGGCGCATTCATCGCGGTTGCGCTTAAATTCGCTTATC	2760	Qy	3781	GGACCAATAAGGACCTCGCGCAACGATCCTGCGGACATTTGGCGGAGAGCAACATTTGT	3840
Db	2701	GAGGACCACTGGCTGATGGGCGCATTCATCGCGGTTGCGCTTAAATTCGCTTATC	2760	Db	3781	GGACCAATAAGGACCTCGCGCAACGATCCTGCGGACATTTGGCGGAGAGCAACATTTGT	3840
Qy	2761	GACGAGCCCGGACGCGTTTACACACCCCATTTACCGGATCCCATGATCAGCCCTGATGGT	2820	Qy	3841	CGCGCGCGCACACTGTGCAACCGTTTACGCTCGTGTCTAAAGACACCAAGGATGTGA	3900
Db	2761	GACGAGCCCGGACGCGTTTACACACCCCATTTACCGGATCCCATGATCAGCCCTGATGGT	2820	Db	3841	CGCGCGCGCACACTGTGCAACCGTTTACGCTCGTGTCTAAAGACACCAAGGATGTGA	3900
Qy	2821	CATGGTTGGAACATGTTCTTGGGGCCCCAACGGGAAAAACCTCACCGGTGAGCGGTTCTA	2880	Qy	3901	TGCCCCAAAGTCTGGATGATCCAGATCTGAAGAGCACCTTTGAACTGGCGGATGTT	3960
Db	2821	CATGGTTGGAACATGTTCTTGGGGCCCCAACGGGAAAAACCTCACCGGTGAGCGGTTCTA	2880	Db	3901	TGCCCCAAAGTCTGGATGATCCAGATCTGAAGAGCACCTTTGAACTGGCGGATGTT	3960

```
QY 3961 CCAGATCATCGTCGGGCCAGGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCAAC 4020
Db 3961 CCAGATCATCGTCGGGCCAGGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCAAC 4020
QY 4021 CTCCTAAGACATCGCTGTCTCCACAGAGCAGCTCAAAGATGTTGGTAAACAGCCAA 4080
Db 4021 CTCCTAAGACATCGCTGTCTCCACAGAGCAGCTCAAAGATGTTGGTAAACAGCCAA 4080
QY 4081 CTGGTTCACCGTGTCTGTAAGATATGCGGACATGTTTCGCCGCTGATTCCTCAATCTT 4140
Db 4081 CTGGTTCACCGTGTCTGTAAGATATGCGGACATGTTTCGCCGCTGATTCCTCAATCTT 4140
QY 4141 GGTGGTGGCGGTCTCTCATGCTATCAACATGTTGGTGGCAGGATCTGTTCGG 4200
Db 4141 GGTGGTGGCGGTCTCTCATGCTATCAACATGTTGGTGGCAGGATCTGTTCGG 4200
QY 4201 TCCGCAATCAGTGGTGGATGTTCCCTCAGATCAGCGGTGTTGTGAGATGATCAAGCT 4260
Db 4201 TCCGCAATCAGTGGTGGATGTTCCCTCAGATCAGCGGTGTTGTGAGATGATCAAGCT 4260
QY 4261 GATGGCATCTGGCGCGTTCGCGTCTTGCCAGTGTGGTGGTTCACCGCAACCAAGG 4320
Db 4261 GATGGCATCTGGCGCGTTCGCGTCTTGCCAGTGTGGTGGTTCACCGCAACCAAGG 4320
QY 4321 TTTCCGTGGCAATGATTCCTGGGCGCGGCATTTGGATGGCGATGGTTCCTCCCAACCT 4380
Db 4321 TTTCCGTGGCAATGATTCCTGGGCGCGGCATTTGGATGGCGATGGTTCCTCCCAACCT 4380
QY 4381 GGTAAAGCGCTAGAGTGGCGCGCACCATGACCGGGCGGAAATGCCAATGTGTCCT 4440
Db 4381 GGTAAAGCGCTAGAGTGGCGCGCACCATGACCGGGCGGAAATGCCAATGTGTCCT 4440
QY 4441 GTTTGGTGGATGTTGCTCAAGCTGGTTACCAGGGCACCGTCTTCTGTGCTGGTGT 4500
Db 4441 GTTTGGTGGATGTTGCTCAAGCTGGTTACCAGGGCACCGTCTTCTGTGCTGGTGT 4500
QY 4501 CTCTTCAGATCTGGCAACCATCGAAGTTCCTGACACAGCGCTGCGGACACTGCGAGA 4560
Db 4501 CTCTTCAGATCTGGCAACCATCGAAGTTCCTGACACAGCGCTGCGGACACTGCGAGA 4560
QY 4561 CTTCTGTGATCACCACAGTGTGACTGTGCTGTGCTCACCAGGCTTCTTACGTTTACG 4620
Db 4561 CTTCTGTGATCACCACAGTGTGACTGTGCTGTGCTCACCAGGCTTCTTACGTTTACG 4620
QY 4621 TGGTCCAGCAATCGCGTGGGTGGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 4680
Db 4621 TGGTCCAGCAATCGCGTGGGTGGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 4680
QY 4681 TTTCCGTGTGTCAGTGGCGGCTGCTGTTTTCGCTGTGCTGTGCTGTGCTGTGCTGT 4740
Db 4681 TTTCCGTGTGTCAGTGGCGGCTGCTGTTTTCGCTGTGCTGTGCTGTGCTGTGCTGT 4740
QY 4741 TGGTCTGACCAAGTCTTCCGCAATGAGCTGAGCTGTGCTGTGCTGTGCTGTGCTGT 4800
Db 4741 TGGTCTGACCAAGTCTTCCGCAATGAGCTGAGCTGTGCTGTGCTGTGCTGTGCTGT 4800
QY 4801 CATCTTCGCAACCGCATCCATGCCCATAATCGCAGAGGTGAGCATGTTTGGCAGTGT 4860
Db 4801 CATCTTCGCAACCGCATCCATGCCCATAATCGCAGAGGTGAGCATGTTTGGCAGTGT 4860
QY 4861 CTTCTAGCAGAGAGTGAAGCTCAAGGCTTGCAGGTGTTGAGGTGTTCCCGCTGT 4920
Db 4861 CTTCTAGCAGAGAGTGAAGCTCAAGGCTTGCAGGTGTTGAGGTGTTCCCGCTGT 4920
QY 4921 TCTTGTATATACAGAGCTGCGATCTTGGTGTGAGCTTCCGCTGCGCTGCGCTGCTA 4980
Db 4921 TCTTGTATATACAGAGCTGCGATCTTGGTGTGAGCTTCCGCTGCGCTGCGCTGCTA 4980
QY 4981 CATTTGATATCGGTACCGAGCTATCGGTGGCGCTTTGATTTGACACTCTTTGATATCAAGC 5040
Db 4981 CATTTGATATCGGTACCGAGCTATCGGTGGCGCTTTGATTTGACACTCTTTGATATCAAGC 5040
```

```
QY 5041 AGTTGGCTTGGCGCTCGAGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATATGGT 5100
Db 5041 AGTTGGCTTGGCGCTCGAGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGATATGGT 5100
QY 5101 CATGTTCTTGGTTTGGCGGTAGTTACCTTTGTCATCGCTTCGGCGCAGGATTGCTTA 5160
Db 5101 CATGTTCTTGGTTTGGCGGTAGTTACCTTTGTCATCGCTTCGGCGCAGGATTGCTTA 5160
QY 5161 TGGCCTTTACTTGTTCGCCGAACGGCAGCATTTGATCCAGATGCAACCCCTGCTCCAGT 5220
Db 5161 TGGCCTTTACTTGTTCGCCGAACGGCAGCATTTGATCCAGATGCAACCCCTGCTCCAGT 5220
QY 5221 GCCTGCAGAAACAGCAAGCCGAAGCAGACACCGCAGAAATTTTCAACACGATTCAC 5280
Db 5221 GCCTGCAGAAACAGCAAGCCGAAGCAGACACCGCAGAAATTTTCAACACGATTCAC 5280
QY 5281 CATCATCCAGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTACGATGCGCAT 5340
Db 5281 CATCATCCAGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTACGATGCGCAT 5340
QY 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTGGCATGCTCCCAACCAAGGGGAGCTGGT 5400
Db 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTGGCATGCTCCCAACCAAGGGGAGCTGGT 5400
QY 5401 TTCACAGTGAAGCAAGATCGGTGGCTTCCCATCTGCTCCCATCTGGTCAACGACGG 5460
Db 5401 TTCACAGTGAAGCAAGATCGGTGGCTTCCCATCTGCTCCCATCTGGTCAACGACGG 5460
QY 5461 CACTAAGCTGAGGATGTTCCCAATGTTGATATCTTGCATGCACATTTGTTGCACACCGT 5520
Db 5461 CACTAAGCTGAGGATGTTCCCAATGTTGATATCTTGCATGCACATTTGTTGCACACCGT 5520
QY 5521 AAACCTCAACGGCAGCACTTTAACCCGCTGAAGAACAGGCGATGAAGTCAAAAGCAGG 5580
Db 5521 AAACCTCAACGGCAGCACTTTAACCCGCTGAAGAACAGGCGATGAAGTCAAAAGCAGG 5580
QY 5581 GGAGCTGCTGTGTAATTCGATATGATGCTTAAAGCTGACAGTTATGAGTTAACAC 5640
Db 5581 GGAGCTGCTGTGTAATTCGATATGATGCTTAAAGCTGACAGTTATGAGTTAACAC 5640
QY 5641 GCCGATTGTTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTTACGGTTTGGCGA 5700
Db 5641 GCCGATTGTTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTTACGGTTTGGCGA 5700
QY 5701 AATTGAAGCGGAGCCCAACCTGCTCAAGTCGCAAAAGAACGCGTGCACCAACACC 5760
Db 5701 AATTGAAGCGGAGCCCAACCTGCTCAAGTCGCAAAAGAACGCGTGCACCAACACC 5760
QY 5761 ATAAGTTGAAACCTTGAGTGTTCGCACACAGTTAGACTAGGGGACGTGACTCTACGCAT 5820
Db 5761 ATAAGTTGAAACCTTGAGTGTTCGCACACAGTTAGACTAGGGGACGTGACTCTACGCAT 5820
QY 5821 CTTTGACACCGGTACCGTACGCTTCAGATTTTAAACCTGTTCAACCGATGCTC 5880
Db 5821 CTTTGACACCGGTACCGTACGCTTCAGATTTTAAACCTGTTCAACCGATGCTC 5880
QY 5881 GGTGTACCTGTGTGGTGCACCCCGCAATCTTCAACCGCAATTTGACATGTTTCCGTCAGC 5940
Db 5881 GGTGTACCTGTGTGGTGCACCCCGCAATCTTCAACCGCAATTTGACATGTTTCCGTCAGC 5940
QY 5941 AGTACGTTTGTATATTTTGGCGCGCTGAA 5969
Db 5941 AGTACGTTTGTATATTTTGGCGCGCTGAA 5969
```

RESULT 2  
AAH68533/c  
ID AAH68533 standard; DNA: 349980 BP.  
XX  
AC AAH68533;  
XX  
DT 26-SEP-2001 (first entry)  
XX

DE C glutamicum coding sequence fragment SEQ ID NO: 7058.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159182.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
XX mutation point of a gene, measuring expression of a gene, analysing  
XX expression profile or pattern of a gene and identifying homologous gene

XX Disclosure; SEQ ID NO: 7058; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
XX sequences from the Corynebacterium glutamicum. These  
XX are useful for identifying the mutation point of a gene derived from a  
XX mutant of corynebacterium, measuring expression amount and  
XX analysing the expression profile or expression pattern of a gene derived  
XX from corynebacterium, and identifying a homologue of a gene derived  
XX from corynebacterium. Corynebacterium bacteria are useful for producing  
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX particularly L-lysine. The present invention is a nucleic acid described  
XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from the  
XX European Patent Office.

XX Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

Query Match 94.8%; Score 5656.2; DB 22; Length 349980;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

QY 1 AGTCGGTCGAGCCACCATGATGTTGGTTCACCGAGCTGCGGAGGCTTTCTACATCT 60  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 AGCTCCCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCAGCCGGTGAAGTTGCG 120  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 115525 ACATCGCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCAGCCGGTGAAGTTGCG 115466  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 121 GAACCCATGGTTCCTTGTGGTGGAGGACAGTGGCGGTGAGAGTTTCAAGTG 180  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 115465 GAACCCATGGTTCCTTGTGGTGGAGGACAGTGGCGGTGAGAGTTTCAAGTG 115406  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 181 TCTCAGTCTTTTAAGTTATGTCATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG 240  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 115405 TCTCAGTCTTTTAAGTTATGTCATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG 115346  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 241 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTTAGACATAGACATCAT 300  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 115345 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTTAGACATAGACATCAT 115286  
DB || || || || || || || || || || || || || || || || || || || || ||

QY 301 CGTCTATGCTTGTGGAAGAACCAATACCTCAGAAAGATGGCAAGTGTGCTT 360  
DB || || || || || || || || || || || || || || || || || || || || ||

Db	115285	CCTCCTATGCTTGTGGGAAGGCAATAACCTCAGAAAGATGCGAAGATGTTGTCATT	115226
QY	361	ATCAAGAAATCAGGTCAAGCAGTTAAAAAATTTGAGGAAAGAAATTTTCOCCTCCG	420
Db	115225	ATCAAGAAATCAGGTCAAGCAGTTAAAAAATTTGAGGAAAGAAATTTGTTACCCCCACG	115166
QY	421	GGGTGATTGATGGCTTCTCCAACTCGAAGCGGCATCATCAGCAACTCTCTGAGAAC	480
Db	115165	GGGTGATTGATGGCTTCTCCAACTCGAAGCGGCATCATCAGCAACTCTCTGAGAAC	115106
QY	481	CAGCACCTAAAAACGACGAGATTCCACCCCGAACTCCACGAGTTTCCCGGTTTATTG	540
Db	115105	CAGCACCTAAAAACGACGAGATTCCACCCCGAACTCCACGAGTTTCCCGGTTTATTG	115046
QY	541	ATCTTCATAATCAGGTGGAACGGTGGCGGCTTTCTTACGGGAACGACGAGCAGCGA	600
Db	115045	ATCTTCATAATCAGGTGGAACGGTGGCGGCTTTCTTACGGGAACGACGAGCAGCGA	114986
QY	601	GGAACACCCGCGCAGTATCACCGCAACATGGCAGCAGCTGATGTTGGCAAGCATGTTT	660
Db	114985	GGAATGCCCGCAGTATCACCGCAACATGGCAGCAGCTGATGTTGGCAAGCATGTTT	114926
QY	661	CGCGCGCGCTGACGAGTGGCAGCGAGTGGAAAAAAGCTTATTCCTTGTGTGAGAGG	720
Db	114925	CGCGCGCGCTGACGAGTGGCAGCGAGTGGAAAAAAGCTTATTCCTTGTGTGAGAGG	114866
QY	721	TCTGTGTGCGCGCATTCACCTCGAGGGGCCCTTTCATCAACGATGCGGCTTGTGTC	780
Db	114865	GCTGTGTGCGCGCATTCACCTCGAGGGGCCCTTTCATCAACGATGCGGCTTGTGTC	114806
QY	781	AAAAACCGGATTCATTTTCCCGCAACCCCAACAGATCTTCCCGGCTGATCCATCGG	840
Db	114805	AAAAACCGGATTCATTTTCCCGCAACCCCAACAGATCTTCCCGGCTGATCCATCGG	114746
QY	841	GAAAAGTTGGATCAAAATCGATCAGTAGCAGCGGAAACTGACAATCTTCTGAGCTTC	900
Db	114745	GAAAAGTTGGATCAAAATCGATCAGTAGCAGCGGAAACTGACAATCTTCTGAGCTTC	114686
QY	901	TGATCTCTGCGGAGCGACACATCATTTCTTCTTCCGGCACACTGATGAGATTG	960
Db	114685	TGATCTCTGCGGAGCGACACATCATTTCTTCTTCCGGCACACTGATGAGATTG	114626
QY	961	ATACCACTACCGAGCGCAATTTGGCTTAAAGAGAAAAAATGTGACGGTCAAGCTAGC	1020
Db	114625	ATACCACTACCGAGCGCAATTTGGCTTAAAGAGAAAAAATGTGACGGTCAAGCTAGC	114566
QY	1021	ATTGTTCATGCGATGCTCCGCTGCATCATAGGCTCCCGGAGCGTGGCGCTTTCG	1080
Db	114565	ATTGTTCATGCGATGCTCCGCTGCATCATAGGCTCCCGGAGCGTGGCGCTTTCG	114506
QY	1081	TGCTCGGCACGTGCGGGGAGCGCATATGTTGAGTTGATCGCGACGCGTGCATTGG	1140
Db	114505	TGCTCGGCACGTGCGGGGAGCGCATATGTTGAGTTGATCGCGACGCGTGCATTGG	114446
QY	1141	CGATGGAACGGTGCATCTAGTCTGTTCCAAACAGCCCTTTTTCATCAGGAGCGCATG	1200
Db	114445	CGATGGAACGGTGCATCTAGTCTGTTCCAAACAGCCCTTTTTCATCAGGAGCGCATG	114386
QY	1201	AAGCCCGCGAATGCCAGACGGTGAATACATTTTGGGCGCTTTTGACGTCACCGTACCG	1260
Db	114385	AAGCCCGCGAATGCCAGACGGTGAATACATTTTGGGCGCTTTTGACGTCACCGTACCG	114326
QY	1261	ATGAGTTCGCGCTCTGCGCGATGGCGGCGCATCGCCGGGGGACACGACATAGCGA	1320
Db	114325	ATGAGTTCGCGCTCTGCGCGATGGCGGCGCATCGCCGGGGGACACGACATAGCGA	114266
QY	1321	GTGAGTTCGTCACACGCTGCGCAGGGGTATGACGCTTATCGACGCGCCCTCCACCT	1380
Db	114265	GTGAGTTCGTCACACGCTGCGCAGGGGTATGACGCTTATCGACGCGCCCTCCACCT	114206
QY	1381	CAACCGTCGCGCTAAATTTCTCGGTCCTTGGGGGATCAGCAAAATCGCTAAATCAACCTG	1440
Db	114205	CAACCGTCGCGCTAAATTTCTCGGTCCTTGGGGGATCAGCAAAATCGCTAAATCAACCTG	114146

QY	1441	CRATTTTGTGTCCTTTGACTCAACCGGCCAGGTGCAAAAGTTCATTTAGTCAATCAAG	1500	QY	2519	CCGACCGCCAGCGATTGCAGTGGAGCGACACTGCCCGAGGCTCTTTACCGGATGCATCCT	2578
Db	114145	TAAATTTGTGTCCTTTGACTCAACCGGCCAGTACAAAGTTCATTTAGACCATCAAG	114086	Db	113065	CCGACCGCCAGCGATTGCAGTGGAGCGACACTGCCCGAGGCTCTTTACCGGATGCATCCT	113006
QY	1501	TACTTTAAAGTACAGTAAAGTATCTGAT - TTAAAGGAGTCCCAACCATGGAATCACT	1559	QY	2579	ATGACCTGGATGGATGCTATTCCCGTGGAGCGCTATTACTAGCGGCACACTTAAACTTT	2638
Db	114085	TAAATTTAAATACGAGCAAAACTTTCCTGATAATAAAGAGTCCGACCATGGAATCATC	114026	Db	113005	ATGACCTGGATGGATGCTATTCCCGTGGAGCGCTATTACTAGCGGCACACTTAAACTTT	112946
QY	1560	ATCTGCAAGACGAGCAAGAAGTCGCAAAAGAGTTCGATGCTTAATCGCAACCTTCGCC	1619	QY	2639	TCTACACCGGCAACTTAAATTTGACGGAAAGCGCGCGCCACCCAAAACCTTGTGGAAG	2698
Db	114025	ATCTGCAAGACGAGCAAGAAGTCGCAAAAGAGTTCGATGCTTAATCGCAACCTTCGCC	113966	Db	112945	TCTACACCGGCAACTTAAATTTGACGGCAAGCGCGCGCCACCCAAAACCTTGTGGAAG	112886
QY	1620	AACAAGGGTGAACCTTGGGGCTTGCAACAGAGATCTCACCACTGAGTACCTACCAAGAG	1679	QY	2699	TCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGTTCGCCCTAAAAATCCGCTTA	2758
Db	113965	ACTAAGGGCGGAACCTTGGGGCTTGCAACAGAGATCTCACCACTGAGTACCTACCAAGAG	113906	Db	112885	TCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGTTCGCCCTAAAAATCCGCTTA	112826
QY	1680	CTCATTCGATGATGAAGCTGGGGAAGTGTCAATCAAGAGTGCNAAGCATCTCTGTGG	1739	QY	2759	TCGAGGACCCCGCAGCGGTTTCACACCCCATTTACCGGATPCCCATGATCAGCGCTGATG	2818
Db	113905	CTCATTCGATGATGAAGCTGGGGAAGTGTCAATCAAGAGTGCNAAGCATCTCTGTGG	113846	Db	112825	TCGAGGACCCCGCAGCGGTTTCACACCCCATTTACCGGATPCCCATGATCAGCGCTGATG	112766
QY	1740	GATGAATACGTGGGACTAACCCGTGACGATGAACAGAGTACTTTAAACCATTCGCAAA	1799	QY	2819	GTGATGGTTGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTCCACGGTCCACGGTTC	2878
Db	113845	GATGAATACGTGGGATTAACGCCGACGATGAACAGAGTACTTTAAACCATTCGTAATA	113786	Db	112765	GTGATGGTTGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTCCACGGTCCACGGTTC	112706
QY	1800	GAGTTCACATGACACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCACAC	1859	QY	2879	TATACCGCTCGACAGATCTTGAACACTGGGAATTCCTCGGTGAATCACCTTTGACCTCA	2938
Db	113785	GAGTTCACATGACACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCACAC	113726	Db	112705	TATACCGCTCGACAGATCTTGAACACTGGGAATTCCTCGGTGAATCACCTTTGACCTCA	112646
QY	1860	CTGTATCCATPACCAAGCAGTCGAGAGTATGAGCAAGATCGCTGCACAAATCCGTTGA	1919	QY	2939	GTGATGCAACACTGGTTCCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT	2998
Db	113725	CTGTATCCATPACCAAGCAGTCGAGAGTATGAGCAAGATCGCTGCACAAATCCGTTGA	113666	Db	112645	GTGATGCAACACTGGTTCCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT	112586
QY	1920	GTTCAAACTCTTGGCATCGCGGAAACGG - CACATCGCTTTCATTGAACCATCATCTTCT	1978	QY	2999	CAAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATCTTCT	3058
Db	113665	GTTCAAACTCTTGGCATCGCGGAAACGGCAGATCGCTTCAATGAGCCATCATCTTCT	113606	Db	112585	CAAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATCTTCT	112526
QY	1979	CTGTGAGGACTGCAAGGTCCAGCGCTGCACCTTAAACTCTGAGGACACACTCGA	2038	QY	3059	GTCCACAGAGTTCGACCGATTCACCATGAGTCTACTACGCAAGCTCTCAGCTCTCAGCTCT	3118
Db	113605	CTGTGAGGACTGCAAGGTCCAGCGCTGCACCTTAAACTCTGAGGACACACTCGA	113546	Db	112525	GTCCACAGAGTTCGACCGATTCACCATGAGTCTACTACGCAAGCTCTCAGCTCTCAGCTCT	112465
QY	2039	TTCTTCAACACCATGCAAGAGTTCACACCCAGCGCTCACCCAGGGTTTGGGCACTTTG	2098	QY	3119	CGGATATGCTCGTCACAACTTGAAGAACGACCTTCGCGCTCTTCGAGGATTCAGCG	3178
Db	113545	TTCTTCAACACCATGCAAGAGTTCACACCCAGCGCTCACCCAGGGTTTGGGCACTTTG	113486	Db	112465	CGGATATGCTCGTCGGAAGCTTGAAGAACGACCTTCGCGCTCTTCGAGGATTCAGCG	112406
QY	2099	TCCCGCGGCAAAACATCGTGTGGTGGCACTGGTGAAGAAAGCCACGCCATCCGC	2158	QY	3179	AGCTGGATTCGCGCATGAATTCAGCACCGCAGGTTCAGTGAACCGTTCGATCGCT	3238
Db	113485	TCCCGCGGCAAAACATCGTGTGGTGGCACTGGTGAAGAAAGCCACGCCATCCGC	113426	Db	112405	AGCTGGATTCGCGCATGAATTCAGCACCGCAGGTTCAGTGAACCGTTCGATCGCT	112346
QY	2159	GGAACCTGTGAAGGCCCACTGATGCTTCTTGCACAGTTCATCTGTAGATGACACAC	2218	QY	3239	GGCTCGTGGGTGGATGGGCTGCCCGCGCAGGATGATCACCCCAACAGTTGCCACGAAG	3298
Db	113425	GGAACCTGTGAAGGCCCACTGACCGCCATGTGCCAGGTTCCATCTGCGAGATGACACAC	113366	Db	112345	GGCTCGTGGGTGGATGGGCTGCCCGCGCAGGATGATCACCCCAACAGTTGCCACGAAG	112286
QY	2219	ATGCCACCATCATCTGTTGATGAGCAGCAGTATCCAAAGCTGGAACGCTGATCCTAC	2278	QY	3299	GATGGGTGCACTGCTGACTGTGCCCGCAGCTTCATTTCCGCAACACCGGATCTACC	3358
Db	113365	AATGCCACCATCATCTGTTGATGAGCAGCAGTATCCAAAGCTGGAACGCTGATCCTAC	113306	Db	112285	GATGGGTGCACTGCTGACTGTGCCCGCAGCTTCATTTCCGCAACACCGGATCTATC	112226
QY	2279	CGTCTCATGGAGCAATTAAGCTGCCTAGAACCAAAAGAAAGTACTGTGTGGGGCTA	2338	QY	3359	AAGAGCTCCTTCTCCAGAGGGGAGTCGGGGGTAAATCAGATCTGTATTAGTTCTGAAC	3418
Db	113305	CGTCTCATGGAGCAATTAAGCTGCCTAGAACCAAAAGAAAGTACTGTGTGGGGCTA	113246	Db	112225	AAGAGCTCCTTCTCCAGAGGGGAGTCAGGGGTAATCAGATCTGTATTAGTTCTGAAC	112166
QY	2339	TGCACACAGACTTTCAGTTTGGCGCTTGGTACCATGCTGCTCCGAGGGCAGGC	2398	QY	3419	CTGTCCGAGTAGACATCCGAGGCAATATTTCCTTCGAGTGGGATGGTTCGCTTGTCTG	3478
Db	113245	TGCACACAGACTTTCAGTTTGGCGCTTGGTACCATGCTGCTCCGAGGGCAGGC	113186	Db	112165	CTGTCCGAGTAGACATCCGAGGCAATATTTCCTTCGAGTGGGATGGTTCGCTTGTCTG	112106
QY	2399	TCAATGATCCCAACGGAATGATGCTGATGGAGATACCCCTCCACGCTCTACTACCAAGCAG	2458	QY	3479	TGGATCTGATGTTGATCGCTGCTGAGTGAAGTAAACCTTGGCGAATTAGTGATCGCGG	3538
Db	113185	TCAATGATCCCAACGGAATGATGCTGATGGAGATACCCCTCCACGCTCTACTACCAAGCAG	113126	Db	112105	TGGATCTGATGTTGATCGCTGCTGAGTGAAGTAAACCTTGGCGAATTAGTGATCGCGG	112046
QY	2459	ATCCAGGTTCCCTTCGCAACCAAGCCAGCGCTGGGCTCACACCAAGCCAGCGGTGA	2518	QY	3539	ACGATAATACAGCATTTAGATTAACCTGAGGTGATGACAGGTTTCAATCGCTTTCCGG	3598
Db	113125	ATCCAGGTTCCCTTCGCAACCAAGCCAGCGCTGGGCTCACACCAAGCCAGCGGTGA	113066	Db	112045	ACGATAATACAGCATTTAGATTAACCTGAGGTGATGACAGGTTTCAATCGCTTTCCGG	111987
				QY	3599	GCCTTCAAAGGTGACACTATTGAGAGATAAGTCATATAAAGGCTCTTTTGTGCGCAATT	3658

Db	111986	GCCTTCAAGGTGACACTATTGAGAGATAAGTCATAAAAGGGCTTTTGGCGAATT	111927	Db	110906	GATTTCCGGTGTCCAGTCCGGGTTCTGCTCTCGGTCTGCTTACTCACCAATCGTCATC	110847
Qy	3659	GTACAAATACCTTCGCAAAATCCCTTGATCGGACACAAATAACAGAGTTTAAATGTTTFA	3718	Qy	4739	ACTGCTCTGCACAGTCTCTCCCGCAATTTAGCTGAGCTGTTTCAACCAAGGTTGATCC	4798
Db	111926	GTACAAATACCTTCGCAAAATCCCTTGATCGGACACAAATAACAGAGTTTAAATGTTTFA	111867	Db	110846	ACTGCTCTGCACAGTCTCTCCCGCAATTTAGCTGAGCTGTTTAAACCAAGGTTGATCC	110787
Qy	3719	GCCTTTGAACAAACATTCATGCTGAATATTTTGTCTTCCCGGTTTAAAGAGAAATTC	3778	Qy	4799	TTCATCTTCGCAACCGCATCTATGGCTAATATCGCCAGGGTGGCGATGTTTGGCAGTG	4858
Db	111866	GCCTTTGAACAAACATTCATGCTGAATATTTTGTCTTCCCGGTTTAAAGAGAAATTC	111807	Db	110786	TTCATCTTCGCAACCGCATCTATGGCTAATATCGCCAGGGTGGCGATGTTTGGCAGTG	110727
Qy	3779	ATGGACATAAAGACCTCGCGCAACGCATCTCTCGCGACATTTGGCGGCGAAGACACATTT	3938	Qy	4859	TTCCTTCCTAGCGAAGAGTGAAGAGCTCAAGGGCTTTCAGGTGCTTCAGGTGCTCCGCT	4918
Db	111806	ATGGACATAAAGACCTCGCGCAACGCATCTCTCGCGACATTTGGCGGCGAAGACACATTT	111747	Db	110726	TTCCTTCCTAGCGAAGAGTGAAGAGCTCAAGGGCTTTCAGGTGCTTCAGGTGCTCCGCT	110667
Qy	3839	GTCCCGCCGACACATGTGCAACGCTTTACGGCTCTGCTCAAGACACCAAGGATG	3898	Qy	4919	GTCTTGTGATTTACAGAGCTTCGATCTTCGGTGTGAACCTTCGCCCTCGCTGGCCGCTTC	4978
Db	111746	GTCCCGCCGACACATGTGCAACGCTTTACGGCTCTGCTCAAGACACCAAGGATG	111687	Db	110666	GTCTTGTGATTTACAGAGCTTCGATCTTCGGTGTGAACCTTCGCCCTCGCTGGCCGCTTC	110607
Qy	3899	GATCGGCAAGTCTGGATGATGATCCAGATCTGAAAGGACCTTTGAAACTGGCGCATG	3958	Qy	4979	TACATTTGTATCGGTACCGAGCTATCGGTGGCGCTTTGATGTCACCTCTTTGATATCAAG	5038
Db	111686	GATCGGCAAGTCTGGATGATGATCCAGATCTGAAAGGACCTTTGAAACTGGCGCATG	111627	Db	110606	TTCATCGGTATCGGTACCGAGCTATCGGTGGCGCTTTGATGTCACCTCTTTAATATCAAG	110547
Qy	3959	TTCAGATCATCTCGCGGCGAGCGGATGTGGATCATGTTTCAAGAACTCGATGACGA	4018	Qy	5039	GCAGTTGCGTTGGGGCGCTGCAGGTTCTTTGGGTGTTGTTCTATTGATGCTCCAGATATG	5098
Db	111626	TTCAGATCATCTCGCGGCGAGCGGATGTGGATCATGTTTCAAGAACTCGATGACGA	111567	Db	110546	GCAGTTGCGTTGGGGCGCTGCAGGTTCTTTGGGTGTTGTTCTATTGATGCTCCAGATATG	110487
Qy	4019	ACCTCOAAGACATCGCTGTGTCCAGAGCAGCTCAAGATGTTGTGCTTAACACGCC	4078	Qy	5099	GTATGTTCTTGGTTTGGCGGCTAGTTACCTTTGTCTATCGCATTCGGCGCAGCATTTGCT	5158
Db	111566	ACCTCOAAGACATCGCTGTGTCCAGAGCAGCTCAAGATGTTGTGCTTAACACGCC	111507	Db	110486	GTATGTTCTTGGTTTGGCGGCTAGTTACCTTTGTCTATCGCATTCGGCGCAGCATTTGCT	110427
Qy	4079	AACCTGTTACGCGTCTGTGAAGGATTTGGCGGACATTTTCCTCCCGCTGATTCGAATC	4138	Qy	5159	TATGCGCTTACTTGGTTTCGGCGCAACGCGAGCATTTGATCCAGATGCAACCGCTGCTCCA	5218
Db	111506	AACCTGTTACGCGTCTGTGAAGGATTTGGCGGACATTTTCCTCCCGCTGATTCGAATC	111447	Db	110426	TATGCGCTTACTTGGTTTCGGCGCAACGCGAGCATTTGATCCAGATGCAACCGCTGCTCCA	110367
Qy	4139	TTCGTTGGTGGCGGCTGCTCATGCTATCAACAATGTGTTGTTGCGCAGGATCTGTTTC	4198	Qy	5219	GTGCTTCGAGCAACCAAGCCGAGCAAGCAACCCGAGCAATTTTCAAAAGCATTTCC	5278
Db	111446	TTCGTTGGTGGCGGCTGCTCATGCTATCAACAATGTGTTGTTGCGCAGGATCTGTTTC	111387	Db	110366	GTGCTTCGAGCAACCAAGCCGAGCAAGCAACCCGAGCAATTTTCAAAAGCATTTCC	110307
Qy	4199	GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTTTGTCTGAGATGATCAAC	4258	Qy	5279	ACCATCATCCAGGCACTTTGACCGGTTGACCGGTGAAGCTATTGACCTGAGCAGCGTCAAGG	5338
Db	111386	GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTTTGTCTGAGATGATCAAC	111327	Db	110306	ACCATCATCCAGGCACTTTGACCGGTTGACCGGTGAAGCTATTGACCTGAGCAGCGTCAAGG	110247
Qy	4259	CTGATGSCATCTCGCGGTTTCGGGTTCTTGCCAGTGTGTTGGTTTCAACGCAACCAAG	4318	Qy	5339	ATGTTTGGCAGCGGAAAGCTTTGGCTCAGGTGTTGCGATCGTCCCGCAACCAAGGCGAGTG	5398
Db	111326	CTGATGSCATCTCGCGGTTTCGGGTTCTTGCCAGTGTGTTGGTTTCAACGCAACCAAG	111267	Db	110246	ATGTTTGGCAGCGGAAAGCTTTGGCTCAGGTGTTGCGATCGTCCCGCAACCAAGGCGAGTTA	110187
Qy	4319	CGTTTCGGTGGAATGAGTTCCTGGGCGCGGATTTGGTATGCGGATGGTTTCCCAAC	4378	Qy	5399	GTCTTCCAGCAGTACGGGAAAGATCGTGTGGCTTCCCATCTGCTCAGCGCTTTGCGAGTC	5458
Db	111266	CGTTTCGGTGGAATGAGTTCCTGGGCGCGGATTTGGTATGCGGATGGTTTCCCGAGC	111207	Db	110186	GTCTTCCCGGTGAGTGAAGATTTGTGTGGCATTTCCCATCTGCGCATTTGCGCATTT	110127
Qy	4379	CTGGTTAAACGGCTACACGCTGCGCGCCACCATGACCGCGGCGAAATGCCAATGTGGTCC	4438	Qy	5459	CGCATTAAGGCTGAGGATGGTTCCCAATGTGGATATCTTTGATGTCACATTTGTTTGCACACC	5518
Db	111206	CTGGTTAAACGGCTACACGCTGCGCGCCACCATGACCGCGGCGAAATGCCAATGTGGTCC	111147	Db	110126	CGCATTAAGGCTGAGGATGGTTCCCAATGTGGATATCTTTGATGTCACATTTGTTTGCACACA	110067
Qy	4439	CTGTTTGGTTTGGATGTTGCTCAAGCTGTTACAGGCGCACCGTGTCTCTGCTGCTGGTG	4498	Qy	5519	GTAAACCTCAACGGCACGCACTTTAACCCGTTGAAGAAGCGGGGATGAAGTCAAGCA	5578
Db	111146	CTGTTTGGTTTGGATGTTGCTCAAGCTGTTACAGGCGCACCGTGTCTCTGCTGCTGGTG	111087	Db	110066	GTAAACCTCAACGGCACGCACTTTAACCCGTTGAAGAAGCGGGGATGAAGTCAAGCA	110007
Qy	4499	GTCTCTTGGATTTCTGCAACGATCGAGAGTTCCTGCAAGCGACTCATGGCAGTCA	4558	Qy	5579	GGGGAGCTGCTGTGTGAATTCGATATTTGATGCCATTTAAGGCTGAGGTTTATGAGTAACC	5638
Db	111086	GTCTCTTGGATTTCTGCAACGATCGAGAGTTCCTGCAAGCGACTCATGGCAGTCA	111027	Db	110006	GGGGAGCTGCTGTGTGAATTCGATATTTGATGCCATTTAAGGCTGAGGTTTATGAGTAACC	109947
Qy	4559	GACTTCTGATCAACCCAGTGTGACTCTGCTCAGCGGCTTCCTTACGTTTCAATTTGCT	4618	Qy	5639	ACCGCGATTTGTTTTCGAAATTTACAGAAACCCGACCTGTAAACACTTTACGGTTTGGGC	5698
Db	111026	GACTTCTGATCACTCCAGTGTGAGCTTGTCTGCTCAGCGGATTTCTTACATTCATCGCC	110967	Db	109946	ACCGCGATTTGTTTTCGAAATTTACAGAAACCCGACCTGTAAACACTTTACGGTTTGGGC	109887
Qy	4619	ATTGGTCCACATGCGCTGGTGGGTGACTTCTGCGCACACCGGCTGAGGAGCTCTAT	4678	Qy	5699	GAAATTTGAAGCGGGAGCAACCTGCTCAACGTCGCAAGAAAGAGCGGTCGCAACACA	5758
Db	110966	ATTGGTCCACATGCGCTGGTGGGTGACTTCTGCGCACACCGGCTGAGGAGCTCTAT	110907	Db	109886	GAAATTTGAAGCGGGAGCAACCTGCTCAACGTCGCAAGAAAGAGCGGTCGCAACACA	109827
Qy	4679	GATTTCCGTTGGTTCAGTCGCGGCTGCTTTTCGGTCTGCTTCTGCTTACTTCAACATTCGTTATC	4738	Qy	5759	CCATAAGTTGAACCTTTCAGTGTTCGACACAGGTTAGCTAGGAGCGTCTGACTTACGC	5818
				Db	109826	CCATAAGTTGAACCTTTCAGTGTTCGACACAGGTTAGCTAGGAGCGTCTGACTTCTACGC	109767

QY 5819 ATCTTTGACACCGGTACCGCTACGCTTCGAGATTTTAAACCTGTTCAACACGAGTCATGCC 5878  
 |||||  
 Db 109766 ATCTTTGACACCGGTACCGCTACGCTTCGAGATTTTAAACCTGTTCAACACGAGTCATGCC 109707  
 |||||  
 QY 5879 TCGGTGCTACTGTGTGCGCCACCGCGCAATCTTCACCCACATGGACATCTTCGTCA 5938  
 |||||  
 Db 109706 TCGGTGCTACTGTGTGCGCCACCGCGCAATCTTCACCCACATGGACATCTTCGTCA 109647  
 |||||  
 QY 5939 GCAGTAGCGTTGATATTTGGCGCGCTG 5967  
 |||||  
 Db 109646 GCAGTAGCGTTGATATTTGGCGCGCTG 109618  
 |||||

## RESULT 3

AAT34177

ID AAT34177 standard; DNA; 6911 BP.

XX AC

XX AAT34177;

XX 22-OCT-1996 (first entry)

XX Coryneform sucrose gene.

XX

XX Sucrose; coryneform; L-amino acid; L-lysine; L-glutamic acid;

XX nucleic acid; sucrose fermentation; ss.

XX

XX Brevibacterium lactofermentum strain ATCC 13869.

XX

XX Key

XX Location/Qualifiers

XX 342..1508

XX /tag= a

XX /label= ORF-F1

XX /product= product shows 24% homology to E. coli

XX N-acetylglucosamine-6-phosphate

XX deacetylase

XX

XX CDS

XX 2338..3612

XX /tag= b

XX /label= ORF-F2

XX /product= sucrose

XX 4438..5358

XX /tag= c

XX /label= ORF-F3

XX /product= product shows 36% homology to B. subtilis

XX Up-N-acetylmuramoylalanyl-D-glutamyl-

XX meso-6-diaminopimelate synthetase

XX

XX CDS

XX 5570..6580

XX /tag= d

XX /label= ORF-F2

XX /product= product shows 39% homology to B. subtilis

XX phospho-N-acetylmuramoylpentapeptide

XX transferase

XX

XX EP724017-A2.

XX

XX 31-JUL-1996.

XX

XX 29-JAN-1996;

XX 96EP-0101228.

XX

XX 30-JAN-1995;

XX 95JP-0012361.

XX

XX (AJIN ) AJINOMOTO KK.

XX (AJIN ) AJINOMOTO CO INC.

XX

XX Hiroshi M, Nagase K, Nakamatsu T, Otsuna S, Sugimoto M;

XX Tsuchiya M, Yasuhiko Y, Matsui H, Yoshihara Y;

XX

XX WPI; 1996-343532/35.

XX

XX Sucrose gene from Coryneform bacteria - used to prepare recombinant

XX microorganisms for improved prodn. of L-amino acids and nucleic

XX acids by fermentation of sucrose

XX

XX

XX

XX

XX

PS Claim 2; Page 20-24; 24pp; English.

XX A DNA fragment (AAT34177) includes an open reading frame (ORF-F2)  
 CC coding for sucrose. It was obt'd. by screening a genomic library  
 CC of Brevibacterium lactofermentum ATCC 13869 with probes (see also  
 CC AAT34178-80) based on homologous regions of known sucrose, levanase  
 CC and invertase genes. An isolated fragment was inserted into pSAC4,  
 CC giving plasmid pSSM30. Escherichia coli JM109/pSSM30 was named  
 CC AJ13047 and deposited as FERM BP-4800. Over-expression of the  
 CC sucrose gene in a coryneform bacterium improves the ability of the  
 CC host to assimilate sucrose, resulting in increased fermentation  
 CC rates and improved productivity of L-amino acids (esp. L-lysine and  
 CC L-glutamic acid) and nucleic acids from raw materials contg.  
 CC sucrose.

XX Sequence 6911 BP; 1575 A; 1784 C; 1892 G; 1660 T; 0 other;

Query Match 61.8%; Score 3687; DB 17; Length 6911;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTCGCTGACGCCACCATTCATGCTGGTGCACCGAGCTTGGCGAGGCTTCTACATCT 60  
 |||||  
 Db 1 AGTCGCTGACGCCACCATTCATGCTGGTGCACCGAGCTTGGCGAGGCTTCTACATCT 60  
 |||||  
 QY 61 ACGCTCCGCTCGCGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTGGCG 120  
 |||||  
 Db 61 ACGCTCCGCTCGCGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTGGCG 120  
 |||||  
 QY 121 GAACCCATGGTTCCTTGTGGTGGAGGACGAGTCCGGTGCAGAGTTCCTCAAGTG 180  
 |||||  
 Db 121 GAACCCATGGTTCCTTGTGGTGGAGGACGAGTCCGGTGCAGAGTTCCTCAAGTG 180  
 |||||  
 QY 181 TCTGAGTTCCTTAAAGTTCATGCTAGTTCGTAAGGCTGAGTAACTCAGTAGAGCTG 240  
 |||||  
 Db 181 TCTGAGTTCCTTAAAGTTCATGCTAGTTCGTAAGGCTGAGTAACTCAGTAGAGCTG 240  
 |||||  
 QY 241 CAACAGCAGGCTCAAGTCCGAAGATTAATTAACCTAGATCCGTAGACATAGACATCATA 300  
 |||||  
 Db 241 CAACAGCAGGCTCAAGTCCGAAGATTAATTAACCTAGATCCGTAGACATAGACATCATA 300  
 |||||  
 QY 301 CGTCTATGCTTGTGTAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGATT 360  
 |||||  
 Db 301 CGTCTATGCTTGTGTAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGATT 360  
 |||||  
 QY 361 ATCAGAAATGTCAGGTCAAGTCAAGTAAATAAATGAGGGAAGATTTTCCCGCCCTCG 420  
 |||||  
 Db 361 ATCAGAAATGTCAGGTCAAGTCAAGTAAATAAATGAGGGAAGATTTTCCCGCCCTCG 420  
 |||||  
 QY 421 GGGTGATTGATGGCTTCTCCAACTCGAAACGGCATCATCACGGAACCTCTCTGAGAAC 480  
 |||||  
 Db 421 GGGTGATTGATGGCTTCTCCAACTCGAAACGGCATCATCACGGAACCTCTCTGAGAAC 480  
 |||||  
 QY 481 CAGCACCTAAAACGCGAGGATTCACCCCGAACCTCCCGACCATTTCCCGGTTTATTG 540  
 |||||  
 Db 481 CAGCACCTAAAACGCGAGGATTCACCCCGAACCTCCCGACCATTTCCCGGTTTATTG 540  
 |||||  
 QY 541 ATCTTCATAATCACGGTGGAAACGGTGGCGGCTTTCCTACGGGAACCGACGACCGCA 600  
 |||||  
 Db 541 ATCTTCATAATCACGGTGGAAACGGTGGCGGCTTTCCTACGGGAACCGACGACCGCA 600  
 |||||  
 QY 601 GGAACACCGCGAGTATCACCGCGNACATGCGCAGCAGCGTATGTCCTCAGAGCATGTTT 660  
 |||||  
 Db 601 GGAACACCGCGAGTATCACCGCGNACATGCGCAGCAGCGTATGTCCTCAGAGCATGTTT 660  
 |||||  
 QY 661 CGGCGCGGCTGACCACTGGCAGCGCAGGTGGAAACCTTATCCCTTGTGTGAAGAGG 720  
 |||||  
 Db 661 CGGCGCGGCTGACCACTGGCAGCGCAGGTGGAAACCTTATCCCTTGTGTGAAGAGG 720  
 |||||  
 QY 721 TCCCTGCTGGCGCATTCACCTCGAGGCGCCCTTTTCATCAACGCATCCCGTGTGGTGCTC 780  
 |||||  
 Db 721 TCCCTGCTGGCGCATTCACCTCGAGGCGCCCTTTTCATCAACGCATCCCGTGTGGTGCTC 780  
 |||||

QY	781	AAAACCCGGATTTCATTTTCCGGGCAACCCACAGATCTTGC	CCGGGTGATCCATGCCG	840
Db	781	AAAACCCGGATTTCATTTTCCGGGCAACCCACAGATCTTGC	CCGGGTGATCCATGCCG	840
QY	841	GAAGGTTGGATCAAAATCGATCACAGTAGCGCGGAAACTG	ACAAATCTTTCTGACGCTC	900
Db	841	GAAGGTTGGATCAAAATCGATCACAGTAGCGCGGAAACTG	ACAAATCTTTCTGACGCTC	900
QY	901	TCGATCTTCGGCAGCGCACCATCATTTGCTTCCTTCGGG	CACATCGATGAGATTTTG	960
Db	901	TCGATCTTCGGCAGCGCACCATCATTTGCTTCCTTCGGG	CACATCGATGAGATTTTG	960
QY	961	ATACCACTACCAAGCAATTCCTTCGCTTAAAGAGAAAATG	TGACGGTACGGCTACGC	1020
Db	961	ATACCACTACCAAGCAATTCCTTCGCTTAAAGAGAAAATG	TGACGGTACGGCTACGC	1020
QY	1021	ATTGTTCAAATGCGATGCTTCGCTGCATATAGGGCTCCC	GGCAGCGTGGGCGCTTTC	1080
Db	1021	ATTGTTCAAATGCGATGCTTCGCTGCATATAGGGCTCCC	GGCAGCGTGGGCGCTTTC	1080
QY	1081	TTGCTGGGACGTCGGGGAGCGATATGTTGAGTTGATCG	CGGAGCGGTGCATTTGG	1140
Db	1081	TTGCTGGGACGTCGGGGAGCGATATGTTGAGTTGATCG	CGGAGCGGTGCATTTGG	1140
QY	1141	CCGATGGAAGCGTCGATCTAGCTCGTTCCAACAACGCC	TTTTTCATCAGCGAGCCCATG	1200
Db	1141	CCGATGGAAGCGTCGATCTAGCTCGTTCCAACAACGCC	TTTTTCATCAGCGAGCCCATG	1200
QY	1201	AAGCCGGCAATGCCACAGCGTAGTACATTTTGGCGT	TTTGAAGTACCGTCAACG	1260
Db	1201	AAGCCGGCAATGCCACAGCGTAGTACATTTTGGCGT	TTTGAAGTACCGTCAACG	1260
QY	1261	ATGAGTCGCGGCTCTCGCGGATGGCGGCCCATCGCC	CGGAGCACAGCACTAGCGA	1320
Db	1261	ATGAGTCGCGGCTCTCGCGGATGGCGGCCCATCGCC	CGGAGCACAGCACTAGCGA	1320
QY	1321	GTGAGTTGTCACACGTCGCGAGGGGTATGAGGTTATPC	AGCGAGCAACCTCCACACT	1380
Db	1321	GTGAGTTGTCACACGTCGCGAGGGGTATGAGGTTATPC	AGCGAGCAACCTCCACACT	1380
QY	1381	CAACGTCGCGGCTAAAATTCGCTTCGGCGATCAGGAA	ATCGCTAAATCAACCCCTG	1440
Db	1381	CAACGTCGCGGCTAAAATTCGCTTCGGCGATCAGGAA	ATCGCTAAATCAACCCCTG	1440
QY	1441	CAAAATTTGTGCTTTTGACTCAAAAGCCGAGTGC	AAAAGTCCATTTAGGTCATCAAG	1500
Db	1441	CAAAATTTGTGCTTTTGACTCAAAAGCCGAGTGC	AAAAGTCCATTTAGGTCATCAAG	1500
QY	1501	TACTTTAAGTACAGTAAACTCTCTGATTTTAAAGG	AGTCCACCATGGAATCACTA	1560
Db	1501	TACTTTAAGTACAGTAAACTCTCTGATTTTAAAGG	AGTCCACCATGGAATCACTA	1560
QY	1561	TCTGCAAGACGAGCAAGAGTTCGGCAAAACAGCTTG	CAGTCCCTAATCGCACCCCTCGCCA	1620
Db	1561	TCTGCAAGACGAGCAAGAGTTCGGCAAAACAGCTTG	CAGTCCCTAATCGCACCCCTCGCCA	1620
QY	1621	ACAAGGTTGAACCTTTGGGCTTGCAACAGGATCCTC	ACACTGAGTACCTACCAAGC	1680
Db	1621	ACAAGGTTGAACCTTTGGGCTTGCAACAGGATCCTC	ACACTGAGTACCTACCAAGC	1680
QY	1681	TCATTCGATGATGAGCTGGGAGTGCATTCGAAG	ACTCAAGGCAATTCCTGTGG	1740
Db	1681	TCATTCGATGATGAGCTGGGAGTGCATTCGAAG	ACTCAAGGCAATTCCTGTGG	1740
QY	1741	ATGAATACGTGGGCAATACCCGTGAGCATGAAAC	AGCTACTTTAAACCATTCGCAAG	1800
Db	1741	ATGAATACGTGGGCAATACCCGTGAGCATGAAAC	AGCTACTTTAAACCATTCGCAAG	1800
QY	1801	AGTTCAGTACCACTGCATCGTTTCATGAAGAGTCTC	ACGCCAGATGGTGCACAC	1860
Db	1801	AGTTCAGTACCACTGCATCGTTTCATGAAGAGTCTC	ACGCCAGATGGTGCACAC	1860
QY	1861	CTGATCCATACGAAGCAGCTGCAGAGTATGAGCA	AAAGATPCGCTGCAGAAATCCGCTGAAG	1920

Db	1861	CTGATCCATCGAAGCAGCTGCGAGATGATGAGGCAAGATCGCTGCGAATCCGTTGAAG	1920
Qy	1921	TTCAAAATCCTTTGGCATCGGGGAAACGGCACATCGCTTTTCATTGAACCATCATCTTCTCT	1980
Db	1921	TTCAAAATCCTTTGGCATCGGGGAAACGGCACATCGCTTTTCATTGAACCATCATCTTCTCT	1980
Qy	1981	GTCAGGACTGACAAAGTCCAGGGCTGCGACCTTAAACHTGTGGAGGACAAACGCTCGATT	2040
Db	1981	GTCAGGACTGACAAAGTCCAGGGCTGCGACCTTAAACHTGTGGAGGACAAACGCTCGATT	2040
Qy	2041	CTTCAACACCATCAAGAGGTCCCAACCCACGCGCTCACCCAGGGTTTGGCACTTTGTC	2100
Db	2041	CTTCAACACCATCAAGAGGTCCCAACCCACGCGCTCACCCAGGGTTTGGCACTTTGTC	2100
Qy	2101	CCGGCGGCAAAACATCTGTTGTGTGGCAACTGTGTGAAGGAAAGCCGACGCGATCCGCGG	2160
Db	2101	CCGGCGGCAAAACATCTGTTGTGTGGCAACTGTGTGAAGGAAAGCCGACGCGATCCGCGG	2160
Qy	2161	AACHTGTGAAGGCCAGTGACTGCTTCTTGCCAGGTTCCATCCTGTAGATGCAACAAT	2220
Db	2161	AACHTGTGAAGGCCAGTGACTGCTTCTTGCCAGGTTCCATCCTGTAGATGCAACAAT	2220
Qy	2221	GCCACCATCATCTGTTGGATGAAGCAGCAGTAGTATCCAAAGCTGGAACGCTGATCACTACCG	2280
Db	2221	GCCACCATCATCTGTTGGATGAAGCAGCAGTAGTATCCAAAGCTGGAACGCTGATCACTACCG	2280
Qy	2281	TCTCATGGAGCAATTAAGCTGCGCTAGAAACAAAGGAAGTACTGTGTGGGGCTATG	2340
Db	2281	TCTCATGGAGCAATTAAGCTGCGCTAGAAACAAAGGAAGTACTGTGTGGGGCTATG	2340
Qy	2341	CACACAGAACTTTCAGATTTTCGGCCCTGCGTACCATGTGACTTCCTCGCAGGCGAGGCTC	2400
Db	2341	CACACAGAACTTTCAGATTTTCGGCCCTGCGTACCATGTGACTTCCTCGCAGGCGAGGCTC	2400
Qy	2401	AATGATCCCAACGAAATGACTGTCATGAGAGATACCTTCCACGCTCTACTACCAGCAGAT	2460
Db	2401	AATGATCCCAACGAAATGACTGTCATGAGAGATACCTTCCACGCTCTACTACCAGCAGAT	2460
Qy	2461	CCAGTTTCCCTTCGCAACAAAGCGCACCGGCTGGGCTCACACCAACAGCCCTTGACC	2520
Db	2461	CCAGTTTCCCTTCGCAACAAAGCGCACCGGCTGGGCTCACACCAACAGCCCTTGACC	2520
Qy	2521	GGACCGCAGCATTCAGTGACGCACTGCCCCAGCGCTCTTTACCGGATGCAATCTAT	2580
Db	2521	GGACCGCAGCATTCAGTGACGCACTGCCCCAGCGCTCTTTACCGGATGCAATCTAT	2580
Qy	2581	GACCTGGATGATGCTATTCGCGTGGAGCCGTATTTACTGACGCACACTTAAACTTTTC	2640
Db	2581	GACCTGGATGATGCTATTCGCGTGGAGCCGTATTTACTGACGCACACTTAAACTTTTC	2640
Qy	2641	TACACCGGCAACCTAAAAATTTGACGGAAGCGCGCGCCACCCAAAACCTTGTGCGAAGTC	2700
Db	2641	TACACCGGCAACCTAAAAATTTGACGGAAGCGCGCGCCACCCAAAACCTTGTGCGAAGTC	2700
Qy	2701	GAGGACCACTGGGCTGATGGCGGCAATTCATCGCGTTTCGCTCAAAAATCCGCTTATC	2760
Db	2701	GAGGACCACTGGGCTGATGGCGGCAATTCATCGCGTTTCGCTCAAAAATCCGCTTATC	2760
Qy	2761	GACGCAACCGCACGGGTTTCACACCCCATTTACCGCATCCATGATCAGCCCTGATGGT	2820
Db	2761	GACGCAACCGCACGGGTTTCACACCCCATTTACCGCATCCATGATCAGCCCTGATGGT	2820
Qy	2821	GATGGTTGGAACATGGTTCTTTGGGGCCGACGCGAAACCTCACCGGTGACGCGGTTCTA	2880
Db	2821	GATGGTTGGAACATGGTTCTTTGGGGCCGACGCGAAACCTCACCGGTGACGCGGTTCTA	2880
Qy	2881	TACCGCTCGACAGATCTTGAAACTGGGAATTCCTCGGTGAATCACTTTGACCTCAGT	2940
Db	2881	TACCGCTCGACAGATCTTGAAACTGGGAATTCCTCGGTGAATCACTTTGACCTCAGT	2940
Qy	2941	GATGCAACACTGGTTCGTCTCTGATCTCGTTCCCGATGCGTACATGTGGGAATGCCCC	3000

Db 2941 GATGCACAACTGGTTCTCTCCTGATCTCGTCCGATGGCTACATGTGGGAATGCCCC 3000  
Qy 3001 AACCTTTTACGTTCCGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTCTGT 3060  
Db 3001 AACCTTTTACGTTCCGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTCTGT 3060  
Qy 3061 CCACAGGATGACCAATACACGATGACGATGAGTTACTACTAGCAAGCTCTGACAGTGC 3120  
Db 3061 CCACAGGATGACCAATACACGATGAGTTACTACTAGCAAGCTCTGACAGTGC 3120  
Qy 3121 GGATATGCTGTCACAAAGCTTGAAGAACGACCTCCGCGTCTTGGAGGATTCAGCGAG 3180  
Db 3121 GGATATGCTGTCACAAAGCTTGAAGAACGACCTCCGCGTCTTGGAGGATTCAGCGAG 3180  
Qy 3181 CTGGATTTCGGCCATGAATCTACGACCGCAGGTTGCAGTAACGGTTCTGATGCTGG 3240  
Db 3181 CTGGATTTCGGCCATGAATCTACGACCGCAGGTTGCAGTAACGGTTCTGATGCTGG 3240  
Qy 3241 CTGCTGGCTGGATGGGCTGCCGCGCAGGATGATCACCCACAGTTCCACAGGAAGA 3300  
Db 3241 CTGCTGGCTGGATGGGCTGCCGCGCAGGATGATCACCCACAGTTCCACAGGAAGA 3300  
Qy 3301 TGGGTGACGCTGACTGTCGCTCCGCAAGCTTCAATTCGGAACACCGGATCTACCAA 3360  
Db 3301 TGGGTGACGCTGACTGTCGCTCCGCAAGCTTCAATTCGGAACACCGGATCTACCAA 3360  
Qy 3361 GAGCTCCTTCCACAGGGGGAGTCGGGGTAATCAGATCTGTATTAGGTTCTGAACCT 3420  
Db 3361 GAGCTCCTTCCACAGGGGGAGTCGGGGTAATCAGATCTGTATTAGGTTCTGAACCT 3420  
Qy 3421 GTCGAGTAGACATCCGAGGCAATATTCCTCGAGTGGGATGCTCCGTTTGTCTGTG 3480  
Db 3421 GTCGAGTAGACATCCGAGGCAATATTCCTCGAGTGGGATGCTCCGTTTGTCTGTG 3480  
Qy 3481 GATCGTAGTGGTATGCTGCTGCTGAGTGAAGTAAACCTGGCGAATAGTATCGCGGAC 3540  
Db 3481 GATCGTAGTGGTATGCTGCTGCTGAGTGAAGTAAACCTGGCGAATAGTATCGCGGAC 3540  
Qy 3541 GATAATACACCAATGAGATACTCAGGTGATGGACAGGTTTCATTCGCTTTTCGGGCG 3600  
Db 3541 GATAATACACCAATGAGATACTCAGGTGATGGACAGGTTTCATTCGCTTTTCGGGCG 3600  
Qy 3601 CTTCAAGGTGACACATTTGAGAGATAAGTAAAGGGTCTTTTGTGGCGAATTGT 3660  
Db 3601 CTTCAAGGTGACACATTTGAGAGATAAGTAAAGGGTCTTTTGTGGCGAATTGT 3660  
Qy 3661 ACAAATCTTCGCAAAATCCCTTGATC 3687  
Db 3661 ACAAATCTTCGCAAAATCCCTTGATC 3687

## RESULT 4

AAH67869  
ID AAH67869 standard; DNA; 1983 BP.  
XX  
AC AAH67869;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 2904.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EF1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR WPI; 2001-376931/40.  
DR P-PSDB; AAG92650.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
XX  
XX Claim 8; SEQ ID NO: 2904; 246pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium, and identifying a homologue of a gene derived  
CC from corynebacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
XX European Patent Office.  
XX  
SQ Sequence 1983 BP; 410 A; 508 C; 558 G; 507 T; 0 other;  
Query Match 31.4%; Score 1874.2; DB 22; Length 1983;  
Best Local Similarity 96.6%; Pred. No. 0;  
Matches 1915; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
Qy 3779 ATGGACCATAGAGGACCTCGCGAAGCATCTCGCGACATTCGCGGAGACAACTT 3838  
Db 1 ATGGACCATAGAGGACCTCGCGAAGCATCTCGCGACATTCGCGGAGACAACTT 60  
Qy 3839 GTCGCGCGCGACACTGTGCAACGCTTTACGCTCGTCTCAAGACACCAAGGATGTG 3898  
Db 61 GTCGCGCGCGACACTGTGCAACGCTTTACGCTCGTCTCAAGACACCAAGGATGTG 120  
Qy 3899 GATCGCCAAAGTCTGGATGATCCAGATCTGAAAGCACCTTTGAACTGGCGGATG 3958  
Db 121 GATCGCCAAAGTCTGGATGATCCAGATCTGAAAGCACCTTTGAAAGCGGTGTATG 180  
Qy 3959 TTCCAGATCATCTCGCGCGGAGGATGATGATGATGATGATGATGATGATGATGATG 4018  
Db 181 TTCCAGATCATCTCGCGCGGAGGATGATGATGATGATGATGATGATGATGATGATG 240  
Qy 4019 ACCTCCAAAGACATCGCTGTGTCCACAGACGAGCTCAAGATGTTGGCTTAACAGGCC 4078  
Db 241 ACCTCCAAAGACATCGCTGTGTCCACAGACGAGCTCAAGATGTTGGCTTAACAGGCC 300  
Qy 4079 AACTGTTTCAGCGGCTGTGAGGATGTTGCGGACATTTTCCTCCGCTGATTCCTCAATC 4138  
Db 301 AACTGTTTCAGCGGCTGTGAGGATGTTGCGGACATTTTCCTCCGCTGATTCCTCAATC 360  
Qy 4139 TTGGTTTGGTGGCGGCTGCTCATGGCTATCAACATGTTGGTTGGCGAGGATCTGTC 4198  
Db 361 TTGGTTTGGTGGCGGCTGCTCATGGCTATCAACATGTTGGTTGGCGAGGATCTGTC 420  
Qy 4199 GGTCCGCAATCAGTGGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258  
Db 421 GGTCCGCAATCAGTGGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 480  
Qy 4259 CTGATGGCATCTGCGCGGCTTCGCTTCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTG 4318  
Db 481 CTGATGGCATCTGCGCGGCTTCGCTTCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTG 540

QY	4319	CGTTTCGGTGGCAATGAGTTCTGGCGCCGCAATTGGTATGGCGATGTGTTCACAACC	4378
Db	541	CGTTTCGGCGGCAATGAGTTCTTGGCGCGCGTATTGGTATGGCGATGTGTTCGGAGC	600
QY	4379	CTGGTTAAGCGGTACAGACGTGGCCGCCACCAATGACCGCGGGCGAATGCCAATGGTGCC	4438
Db	601	TTGGTGAACGGCTACAGACGTGGCGGCCACCATGGCTGGCGGGGAAATGCCAATGTGGTCC	660
QY	4439	CTGTTTGGTTTGGATTTGCTCAAGCTGGTTTACCAGGGCACCGTGCTTCTCTGTGTGGTG	4498
Db	661	CTGTTTGGTTTAGATTTGCCCAAGCGGTTACCAAGGGCACCGTGCTTCTCTGTGTGGTG	720
QY	4499	GTCTCTGTGATTTCTGCAACGATCGAGAAGTTCTCTCCACAAGGACTCATGGGCACTGCA	4558
Db	721	GTTCTTTGATTTCTGCAACGATCGAGAAGTTCTCTGCAACGACTCAAGGGCACTGCA	780
QY	4559	GACTTCTGATCAACCCACGTTGACTCTGCTCTCACCGGTTCTCTTACGTTTCATTGCT	4618
Db	781	GACTTCTGATCACTCCAGTGTGCTGCTCACCGGATTCTCTTACATTCATCTCGCC	840
QY	4619	ATTGTPCCAGCAATCGCTGGGTGGTGACTTTGCTGGCACACGGTCTGCAGGCACTCTAT	4678
Db	841	ATTGGCCCAAGCAATCGCTGGGTGGCGGATGTCTGGCACACGGTCTACAGGCACTTAT	900
QY	4679	GATTTTCGGTGGTCCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTCACCAATCGTTATC	4738
Db	901	GATTTTCGGTGGTCCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTCACCAATCGTTATC	960
QY	4739	ACTGGTCTGCACAGTCTCTTCCGCCAATTGAGCTGGAGCTGTTCACCAAGGTTGATCC	4798
Db	961	ACTGGTCTGCACAGTCTCTTCCGCCAATTGAGCTGGAGCTGTTCACCAAGGTTGATCC	1020
QY	4799	TTCACTCTGCAACCCGATCCATGGCCCAATATCGCGCAGGGTGCAGACTGTTTGGCAGTG	4858
Db	1021	TTCACTCTGCAACCCGATCCATGGTCTAATATCGCCAGGGTGGCGATGTTTGGCAGTG	1080
QY	4859	TTCTTCTCTAGCAGAGAGTGAAGAGCTCAAGGGGCTTGCAGGTCTTCAGGTGTCTCCGCT	4918
Db	1081	TTCTTCTCTGCGAAGAGTGAAGAGCTCAAGGGGCTTGCAGGTCTTCAGGTGTCTCCGCT	1140
QY	4919	GTTCCTGGTATTACAGAGCCTCGGATCTTCGGTGTGAACCTTCGCTCGGCTGGCGGTTTC	4978
Db	1141	GTTCCTGGTATTACGAGCCTCGGATCTTCGGTGTGAACCTTCGCTCGGCTGGCGGTTTC	1200
QY	4979	TACATTTGGTATCGGTACCGCAGCTATCGGTGGCGGCTTTGATTCGACTCTTTGATATCAAG	5038
Db	1201	TTCACTCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTCGACTCTTTAATATCAAG	1260
QY	5039	GCAGTTCGGTTGGCGGCTGCAGGTTCCTTGGGTGTGGTTCTATTGATGCTCCAGATATG	5098
Db	1261	GCAGTTCGGTTGGCGGCTGCAGGTTCCTTGGGTGTGGTTCTATTGATGCTCCAGATATG	1320
QY	5099	GTCAATGTTCTGTTTCGGCGGTAGTTACCTTTGTCATCGCATTCGGCGCAGGATGCT	5158
Db	1321	GTCAATGTTCTGTTGTGAGTTGTTACCTTCTTATCGCATTCGGCGCAGGATGCT	1380
QY	5159	TATGGCCTTTACTTGGTTTCGCCGACAGGCGAGCATTTGATCCAGATGCAACCGTGTCCCA	5218
Db	1381	TATGGCCTTTACTTGGTTTCGCCGACAGGCGAGCATTTGATCCAGATGCAACCGTGTCCCA	1440
QY	5219	GTGCTCGAGACGACCAAGCCGAGCAGACGACCCGCGAATTTTCAACAGTATCC	5278
Db	1441	GTGCTCGAGACGACCAAGCCGAGCAGACGACCCGCGAATTTTCAACAGTATCC	1500
QY	5279	ACCATCATCAGGCACTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTCAAGCATGCC	5338
Db	1501	ACCATCATCAGGCACTTTGACCGGTGAAGCTATTCGCACTGAGCAGCGTCAAGCATGCC	1560
QY	5339	ATGTTTCCGACGGAAAGCTTGGCTCAGGTGTTCGGATTCGTCGCCACCAAGGGCAGCTG	5398
Db	1561	ATGTTTCCGACGGAAAGCTTGGCTCAGGTGTTCGGATTCGTCGCCACCAAGGGCAGTTA	1620

Qy	5399	GTTCACCACTGAGCGGAAGATGCTGGTGGCGCTTCCCATCTGGTCACGGCTTTTCGCAGTC	5455
Db	1621	GTTCCTCCGGTGAGTGGAAAGATGTGGTGGCAATCCCATCTGGCCATGCTTTTCGCAGTT	1680
Qy	5459	CGCACTAAGGCTGAGGATGGTCCAAATGTGGATATCTTGATGCACATTTGTTTCGACACC	5518
Db	1681	CGCACCAGGCTGAGGATGGTCCAAATGTGGATATCTTGATGCACATTTGTTTCGACACA	1740
Qy	5519	GTAACCTCAACGGCAGCGCACTTTAATCCCGCTGAAGAACGAGGCGGATGAAGTCAAAACA	5578
Db	1741	GTAACCTCAACGGCAGCGCACTTTAATCCCGCTGAAGAACGAGGCGGATGAAGTCAAAACA	1800
Qy	5579	GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGCAGGTTATGAGGTAACC	5638
Db	1801	GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGCAGGTTATGAGGTAACC	1860
Qy	5639	ACGCCGATTTGTTTCGAATTACAAGAAAAACCGGACCTGTAACACTTTACGGTTTGGGC	5698
Db	1861	ACGCCGATTTGTTTCGAATTACAAGAAAAACCGGACCTGTAACACTTTACGGTTTGGGC	1920
Qy	5699	GAATTTGAAGCGGGAGCCCACTGCTCAAGCTGCGCAAGAAAGAACGGGTGCCAGCAACA	5758
Db	1921	GAATTTGAAGCGGGAGCCCACTGCTCAAGCTGCGCAAGAAAGAACGGGTGCCAGCAACA	1980
Qy	5759	CCA 5761	
Db	1981	CCA 1983	
RESULT 5			
AAF87497			
ID	AAF87497 standard; DNA; 1656 BP.		
XX	AAF87497;		
AC			
XX			
DT	09-JUL-2001 (first entry)		
XX			
DE	Corynebacterium thermoaminogenes scrB nucleotide sequence.		
XX			
KW	Corynebacterium; thermophilic; amino acid biosynthesis; enzyme;		
KW	thermotolerant; aceA; accB; dtsR1; dtsR2; pfk; scrB; gluABCD;		
KW	pdhA; pc; ppc; acn; icd; lpd; odhA; ds.		
XX			
OS	Corynebacterium thermoaminogenes.		
XX			
FH	Key		
FT	Location/Qualifiers		
FT	CDS		
FT	309..1598.		
FT	/*tag= a		
FT	/product= "scrB protein"		
XX			
PN	WO200125447-A1.		
XX			
PD	12-APR-2001.		
XX			
PF	04-OCT-2000; 2000WO-JP06913.		
XX			
PR	04-OCT-1999; 99JP-0282716.		
PR	01-NOV-1999; 99JP-0311147.		
PR	21-APR-2000; 2000JP-0120687.		
XX			
PA	(AJIN ) AJINOMOTO CO INC.		
XX			
PI	Hirano S, Nonaka G, Matsuzaki Y, Akiyoshi N, Nakamura K, Kimura E;		
PI	Osumi T, Matsul K, Kawanara Y, Kuranashi O, Nakamatsu T;		
PI	Sugimoto S;		
XX			
DR	WPI; 2001-300170/31.		
DR	P-PSDB; AAB83190.		
XX			
PT	Proteins and their DNA useful for microbial production of L-amino acids		
PT			
PS	Claim 27; Page 199-201; 215pp; Japanese.		

xx	The present sequence is provided in a specification relating to genes encoding thermophilic amino acid biosynthesis system enzymes of CC the thermo tolerant bacterium Corynebacterium thermautotrophicus.
CC	the novel proteins retain at least 30% isocitrate lyase activity after heating at 50oC for 5 minutes. DNA fragments encoding the CC enzymes were isolated from a Corynebacterium thermautotrophicus chromosomal DNA plasmid library by PCR. The DNA may be used for developing strains of amino acid producing microorganisms.
XX	
SQ	Sequence 1656 BP; 392 A; 456 C; 432 G; 376 T; 0 other;
	Query Match            24.4%; Score 1457; DB 22; Length 1656; Best Local Similarity 93.2%; Pred. No. 0; Matches 1545; Conservative 0; Mismatches 110; Indels 2; Gaps 2;
QY	2031 ACGTTCGATCTTCAACACCATCGAAGAGTGCCAACGCCGCCTCACCCAGGGTTGG 2090 
Dd	1 AGCCCCGATTCTCAACACTATCGAAGAGTCCC AACCCAGCGTTGACCCAGGGCTTG 60
QY	2091 GCATTTTGTCGGCGCGCAAACAATCATGTGTTGPGGAATGTPGAAGAAAGCGAC - 2149 
Dd	61 GTACTTTGTCCGCGCGCAAAATATCATGTGTTGGTAGCTGGCCAAGAAAAAGCAGA 120 
QY	2150 GCCATCCGCGAACTGTGGAAGGCCAGTGACTGCTTCTTGCACAGGTTCCATCTCTAG 2209 
Dd	121 GCCATCCGCGAACTGTGGAAGTCCAGTGACTGCTTCTTGCACAGGTTCCATCTCGAA 180 
QY	2210 ATGCACAACATGCCACCATCATCGTTGGATGTAAGCAGCAGTATCCAGCTGGAAAACGT 2269 
Dd	181 ATGCACAACAGCCACCATCATCGTTGATGAAGCAGCAGCATCCAGCTGAAAAATGCT 240 
QY	2270 GATCACTPACCGTCTCATGAGCAATAAAGCTGGCTAGAAACAAAAAGAAAGTACTGT 2329 
Dd	241 GACCATPACCGTCTCATGAGCAATAAAGCTGGCTAGAAACAAAAAGAAAGTACTGT 300 
QY	2330 GTGGGGCTATGCACACAGAATTTCCAGTTTTGGCCCTCGGTACCATGTGACTCCTCCGC 2389 
Dd	301 GTGGGGCTATGCACACAGAATTTCCAGTTTTGGCCCTCGGTACCATGTGACTCCTCCGC 360 
QY	2390 AGGCGAGCCTCAATGATCCCAAGGAATGTACGTCGATGGAGATACCCTCCACGTCTACT 2449 
Dd	361 AGGCGAGACTCAATGATCCCAATGGAATGTACGTCGATGGAGATACCCTCCACGTCTACT 420 
QY	2450 ACAAGCAGGATCCAGGTTTTCCCTTCCACCAAGACGCCACCGGTGGGGCTCACACACCA 2509 
Dd	421 ACCAGCAGCATCCAGGTTTTCCCTTCCACCAAGACGCCACCGGTGGGGCTCACACACCA 480 
QY	2510 CGCGGTTGACCGGACCGCAGCGATTCAGATGGAGCAGCACTGCCGAGCGCTCTTTACCCGG 2569 
Dd	481 CGCGGTTGACCGGACCGCAGCGATTCAGATGGAGCAGCACTGCCGAGCGCTCTTTACCCGG 540 
QY	2570 ATGCATCCTATGACCTGGATGATGCTATTTCGGTGGAGCGGTATTTACTGACGGCACAC 2629 
Dd	541 ATTGATCCTATGACCTGGATGATGCTATTTCGGGCGGAGCGGTATTTCTTGACGGCACGC 600 
QY	2630 TTAAACTTTTCTACACCGCAACTAAAAATTTACGGAAAGCGCGCGCCACCAAAAAACC 2689 
Dd	601 TTAAACTTTTCTACACCGCAACCGAAAAATTGACGGCAAGCGCGCGCCACCAAAAAACC 660 
QY	2690 TTGTCGAAGTCGAGGACCACTGGGCTGATGGCGGCAATTCATCGCGGTTGCGCTTAAAA 2749 
Dd	661 TCGTTCGAAGTCGAGGACCAACTGGGCTGATGGCGGCAATTCATCGCGGTTGCGCTTAAAA 720 
QY	2750 ATCCGCTTATCGAGGACCGCCAGCGGTTTCACACCCCATTATCCGCGGATCCCATGATCA 2809 
Dd	721 ATCCGCTTATCGAGGACCGCCAGCGGTTTTAGCCCCCACTACCGCGATCCCATGATCA 780 
QY	2810 GGCTGTATGCTGATGTTGNAATGTTCTTGGGGCCCAACGGAAAAA CCTCACCGGTG 2869 
Dd	781 GGCTGTATGGGATGTTGGAAGATGTTCTTGGGGCTCACGGCGAAAAA CCTCACCGGTG 840 
QY	2870 CAGCGGTTCTATACCGCTCGACAGATCTTGAAAACTGGGAATTTCTCGGGTGAATACCT 2929 

```

XX 27-JUN-2000; 2000WO-IB00973.
XX 01-JUL-1999; 99US-0142691.
XX 23-AUG-1999; 99US-0150310.
XX 03-SEP-1999; 99DE-1042095.
XX 03-SEP-1999; 99DE-1042097.
XX (BADI ) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;
XX WPI; 2001-080989/09.
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
XX sugar phosphotransferase system proteins or their portions, useful for
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX for transformation -
XX Claim 3; Page 98-101; 144pp; English.
XX The present invention relates to Corynebacterium glutamicum
XX phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
XX The PTS nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria, the typing or identification of C. glutamicum or
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX and as markers for transformation.
XX SQ Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;

Query Match      23.78; Score 1412.6; DB 22; Length 1527;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1462; Conservative 0; Mismatches 64; Indels 1; Gaps 1;

QY 4259 CTGATGCACTGCGCGGTTCGGTCTTCCAGTGTGGTTCACCGCAACCAAG 4318
DB 1 CTCATGCACTGCGCGGTTCGGTCTTCCAGTGTGGTTCACCGCAACCAAG 60

QY 4319 CGTTTCGGTGCATAGTTCCTCGGGCGCG -CGATTGGTATGCGATGTGTTCGCAAC 4377
DB 61 CGTTTCGGTGCATAGTTCCTCGGGCGCGCGTATGTTGATGCGATGTGTTCGCGAG 120

QY 4378 CCTGTTAAAGGCTACGACGTGGCGCCACCATGACCGCGGGGGAATGCCAATGTGGTC 4437
DB 121 CTTGGTGAACGGCTACGACGTGGCGCCACCATGGCTGCGGGCGGAATGCCAATGTGGTC 180

QY 4438 CCTGTTGGTGTGATGTGCTCAAGCTGTTACCGAGGACCGTGTCTTCTGTGCTGGT 4497
DB 181 CCTGTTGGTGTGATGTGCTCAAGCGGTTACCGAGGACCGTGTCTTCTGTGCTGGT 240

QY 4498 GGTCTCTGGATTCTGCAACGATCGAGAAGTTCCTGCACAAGCGACTCATGGCACTGC 4557
DB 241 GGTTCCTGGATTCTGCAACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGCACTGC 300

QY 4558 AGACTTCCTGATACCCAGTGTGATCTGTGCTGCTCACCGGTTCCTTACGTTCAATGC 4617
DB 301 AGACTTCCTGATCACTCCAGTGTGATCTGTGCTGCTCACCGGATTCCTTACATTCATCGC 360

QY 4618 TATTGTTCCCAATAGGCTGGTGGTGGTGTGCTGTCGACACGCTCTGAGGACTCTA 4677
DB 361 CATTTGGCCCAATAGGCTGGTGGTGGTGTGCTGTCGACACGCTCTACAGGACTCTTA 420

QY 4678 TGATTCGGTGGTCCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTACCAATCGTAT 4737
DB 421 TGATTCGGTGGTCCAGTGGCGGTCTGCTTTTCGGTCTGGTCTACTACCAATCGTAT 480

QY 4738 CACTGGTCTGCACAGTTCCTTCCCGCAATTTAGCTGGAGCTGTTCACCGAGGTGATC 4797
DB 481 CACTGGTCTGCACAGTTCCTTCCCGCAATTTAGCTGGAGCTGTTCACCGAGGTGATC 540

QY 4798 CTTTCATCTTCGCAACCGCATCCATGGCCCAATATCGCGAGGCTGACGATGTTTGGCAGT 4857

```

RESULT 7  
 ABS65346  
 ID - ABS65346 standard; DNA; 1527 BP.  
 XX  
 AC ABS65346;

```

DB 541 CTTTCATCTTCGCAACGGCATCTATGGCTAATAATCCCCAGGCTGCGGATGTTTGGCACT 600
QY 4858 GTTCTTCTCTAGCGAAGAGTGAAGAGCTCAAGGGCCCTTGCAGGTGCTTCAGGTGTCTCCGC 4917
DB 601 GTTCTTCTCTAGCGAAGAGTGAAGAGCTCAAGGGCCCTTGCAGGTGCTTCAGGTGTCTCCGC 660
QY 4918 TGTCTTCTGTATTACAGAGCCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGCGTGCCTGT 4977
DB 661 TGTCTTCTGTATTACAGAGCCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGCGTGCCTGT 720
QY 4978 CTACATTTGATATCGGTACCGCAGCTATCGGTGGCCCTTTGATTTGCACATCTTTTATATCAA 5037
DB 721 CTTTCATCGTATCGGTACCGCAGCTATCGGTGGCCCTTTGATTTGCACATCTTTTATATCAA 780
QY 5038 GCGAGTTCGGTTCGGCGCTGCGAGTTCCTTGGTGTGTTCTTCTATTGATGCTCCAGATAT 5097
DB 781 GCGAGTTCGGTTCGGCGCTGCGAGTTCCTTGGTGTGTTCTTCTATTGATGCTCCAGATAT 840
QY 5098 GGTTCATGTTCTTGGTTCGGCGGTAGTTACCTTTGTCATCGCATTCGCGCGCAGCGATTGC 5157
DB 841 GGTTCATGTTCTTGGTTCGGCGGTAGTTACCTTTGTCATCGCATTCGCGCGCAGCGATTGC 900
QY 5158 TTATGGCCCTTACTTGGTTCGGCGAAGCGCAGCATTCGATCCAGATGCAACCGCTGCTCC 5217
DB 901 TTATGGCCCTTACTTGGTTCGGCGAAGCGCAGCATTCGATCCAGATGCAACCGCTGCTCC 960
QY 5218 AGTGGCTGCGAGGAACCAACCAAGCCGAGCAGAACCCGCGAGAAATTTTCAACGATTC 5277
DB 961 AGTGGCTGCGAGGAACCAACCAAGCCGAGCAGAACCCGCGAGAAATTTTCAACGATTC 1020
QY 5278 CACCATATCCAGGACCTTTGACCGGTGAAGCTATCCACATGACGACGCTCAGCGATGC 5337
DB 1021 CACCATATCCAGGACCTTTGACCGGTGAAGCTATTCGCGTGAAGCTATTCGCGTGAAGCT 1080
QY 5338 CATGTTTGCAGCGAAAGCTTGGTTCAGGTTCGCGATCGTCCCGCACCAAGGGGCACT 5397
DB 1081 CATGTTTGCAGCGAAAGCTTGGTTCGCGGCTTGCATCGTCCCGCACCAAGGGGCACT 1140
QY 5398 GGTTCACAGTACGCGGAAGATCGTGTGGCTTCCCATCTGCTCAGCTTCGCTTCGCACT 5457
DB 1141 AGTTTCTCGGTGAGTGAAGATTTGGTGGCATTCGCCATCTGCGCATGCTTTTCGCACT 1200
QY 5458 CCGCACTAAGGCTGAGGATGTTTCCAAATGTGATATCTTGTATGSCATTTGGTTCGACAC 5517
DB 1201 TCGCACCAAGGCTGAGGATGTTTCCAAATGTGATATCTTGTATGSCATTTGGTTCGACAC 1260
QY 5518 CGTAAACCTCAACGGCACCGACTTTTAAACCGCTGAAGAGCAGGCGATGAGTCAAGC 5577
DB 1261 AGTAAACCTCAACGGCACCGACTTTTAAACCGCTGAAGAGCAGGCGATGAGTCAAGC 1320
QY 5578 AGGGAGCTGCTGTGTGAATTGATTTGATGCCATTAAAGGCTGAGGCTTATGAGTAAAC 5637
DB 1321 AGGGAGCTGCTGTGTGAATTGATTTGATGCCATTAAAGGCTGAGGCTTATGAGTAAAC 1380
QY 5638 CACGCGGATGTTGTTTCGAATTACAGAAACCGGACCTGTAACACACTTACGGTTTGGG 5697
DB 1381 CACGCGGATGTTGTTTCGAATTACAGAAACCGGACCTGTAACACACTTACGGTTTGGG 1440
QY 5698 CGAAATGAAGCGGGAGCCAACTGCTCAACGTCGCAAAAGAAAGAGCGGTGCCAGCAAC 5757
DB 1441 CGAAATGAAGCGGGAGCCAACTGCTCAACGTCGCAAAAGAAAGAGCGGTGCCAGCAAC 1500
QY 5758 ACCATTAAGTTGAACCTTGAAGTTCG 5784
DB 1501 ACCATTAAGTTGAACCTTGAAGTTCG 1527

```

XX 15-NOV-2002 (first entry)  
XX DNA encoding C. glutamicum metabolic pathway (MP) protein #5.  
XX Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;  
KW cofactor; nucleotide; nucleoside; trehalose; fine chemical production;  
KW organic acid; non-proteinogenic amino acid; purine base; carbohydrate;  
KW pyrimidine base; lipid; unsaturated fatty acid; diols; polyketide;  
KW aromatic compound; food industry; animal feed; cosmetic industry;  
KW pharmaceutical industry; gene; ds.  
XX Corynebacterium glutamicum AFCC 13032.  
OS WO200251231-A1.  
XX 04-JUL-2002.  
XX 22-DEC-2000; 2000WO-EF13143.  
XX 22-DEC-2000; 2000WO-EF13143.  
XX (BADI) BASF AG.  
XX Pompejus M, Kroege B, Zelder O, Schroeder H;  
XX WPI; 2002-643289/69.  
XX P-PSDB; ABG80325.  
XX New metabolic pathway genes of Corynebacterium glutamicum for producing  
XX fine chemicals, e.g. lipids, (un)saturated fatty acids, vitamins  
XX cofactors or enzymes used in food, feed, cosmetics or pharmaceutical  
XX industries -  
XX Claim 1; Page 108-111; 176pp; English.  
XX The present invention relates to the isolation of Corynebacterium  
XX glutamicum metabolic pathway (MP) proteins, and the polynucleotide  
XX sequences encoding them. The MP proteins are enzymes involved in  
XX the metabolism of molecules important for the normal functioning  
XX of cells (e.g. amino acids, vitamins, cofactors, nucleotides and  
XX nucleosides, or trehalose). The polynucleotide sequences encoding  
XX the MP proteins are useful for producing fine chemicals, particularly  
XX organic acids, non-proteinogenic amino acids, purine and pyrimidine  
XX bases, nucleosides, nucleotides, lipids, (un)saturated fatty acids,  
XX diols, carbohydrates, aromatic compounds, vitamins, cofactors,  
XX polyketides and enzymes. The fine chemicals are useful in the food,  
XX animal feed, cosmetic or pharmaceutical industries. ABS65342-ABS65364  
XX encode the C. glutamicum MP proteins of the invention.  
XX Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;  
Query Match 23.7%; Score 1412.6; DB 24; Length 1527;  
Best Local Similarity 95.7%; Pred. No. 0;  
Matches 1462; Conservative 0; Mismatches 64; Indels 1; Gaps 1;  
QY 4259 CTGATGGCATCTCGCGGTTCTTCCAGTGTGGTTCACCGCAACCAAG 4318  
DB 1 CTGATGGCATCTCGCGGTTCTTCCAGTGTGGTTCACCGCAACCAAG 60  
QY 4319 CGTTTCGGTGGCAATGATGTTCTTGGGCGCGG-CGATTGATGCGGATGTTCCCAAC 4377  
DB 61 CGTTTCGGGCAATGATGTTCTTGGGCGCGCGGCGGATGTTGCGGATGTTCCCGAG 120  
QY 4378 CTTGGTTAAGGTTACGACGTGGCGCGGCGGACCATACCGGGGGGGAATGCCAATGGTGC 4437  
DB 121 CTTGGTTAAGGTTACGACGTGGCGCGGCGGACCATGCGTGGGGAATGCCAATGGTGC 180  
QY 4438 CCTGTTGGTTGGATGTTCTCAAGTGGTTACGAGGCGCGGTTCTTCTGTTGGT 4497  
DB 181 CCTGTTGGTTGGATGTTCTCAAGTGGTTACGAGGCGCGGTTCTTCTGTTGGT 240  
QY 4498 GGTCTCTGGATTCTGGCAACGATCGAGAAAGTTCCTGCAACAGCGACTCATGGGCACTGC 4557

DB 241 GGTTCCTGGATTCTGGCAACGATCGAGAAAGTTCCTGCAACAGCGACTCAAGGCACTGC 300  
QY 4558 AGACTTCCTGATFACCCAGTGTGTGACTCTGCTGTCTACCGGCTTCCTAGTTCATGC 4617  
DB 301 AGACTTCCTGATFACCTCCAGTGTGACTCTGCTGTCTACCGGATTCCTTACATTCATCGC 360  
QY 4618 TATTGGTCCAGCAATGCTGGTGGGTGACTTGTGTGCAACAGGTTCTGACGAGGACTCA 4677  
DB 361 CATTGGCCCAACGATGCTGGTGGGTGACTGCTGTGCAACAGGTTCTGACGAGGACTCA 420  
QY 4678 TGATTTCGGTGGTCCAGTGGGCGGTCTGTTTTCGGTCTGTCTACTACCAACATGTTAT 4737  
DB 421 TGATTTCGGTGGTCCAGTGGGCGGTCTGCTCTGGTCTGTCTACTACCAACATGCTCAT 480  
QY 4738 CACTGCTCTGCACCAAGTCTTCCCGCCAAATGAGTGGAGCTGTTTCAACAGGTTGATC 4797  
DB 481 CACTGCTCTGCACCAAGTCTTCCCGCCAAATGAGTGGAGCTGTTTCAACAGGTTGATC 540  
QY 4798 CTTTCATCTTCGCAACCGCATCATATGCGGCAATATGCGGAGGTTGACGATGTTTGGCAGT 4857  
DB 541 CTTTCATCTTCGCAACCGCATCATATGCGGCAATATGCGGAGGTTGCGGCAATGTTGGCAGT 600  
QY 4858 GTTCTTCCTAGCGAAGAGTGAAGAGTCAAGGGCTTCAGAGTGTCTCAGTGTCTCCGC 4917  
DB 601 GTTCTTCCTAGCGAAGAGTGAAGAGTCAAGGGCTTCAGAGTGTCTCAGTGTCTCCGC 660  
QY 4918 TGTTCCTTGTATTCAGAGCTTGCATCTTCCGTTGTGAACCTTCGCTTCGCTTCGCTTC 4977  
DB 661 TGTTCCTTGTATTCAGAGCTTGCATCTTCCGTTGTGAACCTTCGCTTCGCTTCGCTTC 720  
QY 4978 CTACATTCGTATCGGTACCGAGCTATCGGTGGGCTTTGATTCGACTCTTTGATATCA 5037  
DB 721 CTTCATCGGTATCGGTACCGAGCTATCGGTGGGCTTTGATTCGACTCTTTTAAATCA 780  
QY 5038 GGCAGTTCGTTGGGCGTTCAGAGTTCCTTGGTGTGTTTCTTCTATTGATGCTCCAGATAT 5097  
DB 781 GGCAGTTCGTTGGGCGTTCAGAGTTCCTTGGTGTGTTTCTTCTATTGATGCTCCAGATAT 840  
QY 5098 GGTCTATGTTCTTGGTTCGCGGTAGTACCTTTGCTCATCGATTCGCGGAGGATTCG 5157  
DB 841 GGTCTATGTTCTTGGTTCGCGGTAGTACCTTTGCTCATCGATTCGCGGAGGATTCG 900  
QY 5158 TTATGGCTTTACTTGGTTCGCGGAGGAGGATTCGATTCAGATGCAACCGTTCGCTCC 5217  
DB 901 TTATGGCTTTACTTGGTTCGCGGAGGAGGATTCGATTCAGATGCAACCGTTCGCTCC 960  
QY 5218 AGTGCCTTCAGGAGCAACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5277  
DB 961 AGTGCCTTCAGGAGCAACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020  
QY 5278 CACCATCATCCAGGCAACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTCAGCGATGC 5337  
DB 1021 CACCATCATCCAGGCAACCTTTGACCGGTGAAGCTATTCGATGAGCAGCGTCAGCGATGC 1080  
QY 5338 CATGTTTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5397  
DB 1081 CATGTTTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140  
QY 5398 GGTTCACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5457  
DB 1141 AGTTCCTCGGTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200  
QY 5458 CCGCACTAAGGCTGAGGATGTTTCAATGTTGATATCTTGTATGATGATGTTGTTGCGAC 5517  
DB 1201 TCGCACCAGGCTGAGGATGTTTCAATGTTGATATCTTGTATGATGATGTTGTTGCGAC 1260  
QY 5518 CGTAAACCTCAACGCGCACTTTAAACCGCTGAAGAGGAGGAGGAGGAGGAGGAGGAGG 5577  
DB 1261 AGTAAACCTCAACGCGCACTTTAAACCGCTGAAGAGGAGGAGGAGGAGGAGGAGG 1320  
QY 5578 AGGGAGGCTGCTGTGTAATTCGATATTCATGCCATTAAAGCTTCAGGTTATGAGGTAAC 5637

1321	AGGGAGCGTGTGTGTGTAATTCGATATGATGCCATTAAGGCTGCAGGCTTAGGTAAAC	1380
Db		
5638	CACGCCGATTGTTGTTTCGAAATTACAAAGAAAACCGGACCTGTAACACTTACGGTTTGGG	5697
QY		
1381	CACGCCGATTGTTGTTTCGAAATTACAAAGAAAACCGGACCTGTAACACTTACGGTTTGGG	1440
Db		
5698	CGAAATTGAAGCGGGAGCCCACTGCTCAACCTGCACGAAGAAAGCGGTGCCAGCAAC	5757
QY		
1411	CGAAATTGAAGCGGGAGCCCACTGCTCAACCTGCACGAAGAAAGCGGTGCCAGCAAC	1500
Db		
5758	ACCATAAAGTTGAACCTTGAGTGTTTCG	5784
QY		
1501	ACCATAAAGTTGAACCTTGAGTGTTTCG	1527
Db		

XX	Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;
XX	WPI; 2001-061975/07.
XX	P-PSDB; AAB79416.
XX	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT	metabolism and oxidative phosphorylation protein for production or
PT	modulation of production of fine chemicals e.g. amino acids,
PT	carbohydrates or enzymes -
XX	Claim 3; Page 627-629; 1246pp; English.
XX	AAP71360 to AAP71750 encode the Corynebacterium glutamicum sugar
CC	metabolism and oxidative phosphorylation (SMP) proteins given in
CC	AAB79423 to AAB 79633 which are involved in carbon metabolism and
CC	energy production. The C. glutamicum SMP gene can be used in vectors
CC	(ii) for expression in host cells and production or modulation of
CC	C. glutamicum, such as, an organic acid, a proteinogenic
CC	or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC	acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC	cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC	(iii) encoded by them are used for diagnosing the presence or activity of
CC	Corynebacterium diptheriae in a subject. (i), (ii), (iii) or host cells
CC	containing them are used to map genomes of organisms related to
CC	C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC	in evolutionary studies, in determining SMP protein regions required
CC	for function, in modulating SMP protein activity, in modulating the
CC	metabolism of sugars, and in modulating high-energy molecule production
CC	in a cell (i.e. ATP, NADPH).
XX	Sequence 1342 BP; 312 A; 373 C; 359 G; 298 T; 0 other;
SQ	
Query Match	22.1%; Score 1321.2; DB 22; Length 1342;
Best Local Similarity	99.0%; Pred. No. 0;
Matches 132;	Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy	2238 ATGAAGCAGCAGTATCCAAGCTGGAAACGCTGATCACTACCGTCTCATGGACAATTAA 2297
Db	1 ATGAAGCAGCAGCATCCAAGCTGGAAACGCTGATCACTACCGTCTCATGGACAATTAA 60
Qy	2298 AGCTGCCGTAGAACAACAAAAGAAAGTACTGTGTGGGCTATGCACACAGAACTTTCCAG 2357
Db	61 AGCTGCCGTAGAACAACAAAAGAAAGTAGTGTTGGGGCTATGCACACAGAACTTTCCAG 120
Qy	2358 TTGTGCGCCCTGCGTGATTCACATGTGACTCTCCGACAGGCAGGCTCAATGATCCCACGGAAT 2417
Db	121 TTGTGCGCCCTGCGTGATTCACATGTGACTCTCCGACAGGCAGGCTCAATGATCCCACGGAAT 180
Qy	2418 GTACGTGCGATGGAGATACCCTCCACGTCTACTACACGACAGATCCAGGTTTCCCTTCGC 2477
Db	181 GTACGTGCGATGGCGATACCCTCCACGTCTACTACACGACAGATCCAGGTTTCCCTTCGC 240
Qy	2478 ACCAAGCGCACCGGCTGGGCTCACACACACCGCGTTGACCGGACCGCAGGATGCA 2537
Db	241 ACCAAGCGCACCGGCTGGGCTCACACACACCGCGTTGACCGGACCGCAGGATGCA 300
Qy	2538 GTGACGCACTGCCCGACGCTCTTTACCCGGATGCATCCTATGACTGATGGATGCTA 2597
Db	301 GTGACGCACTGCCCGACGCTCTTTACCCGGATGCATCCTATGACTGATGGATGCTA 360
Qy	2598 TTCGGTGGAGCGGTATTTACTACGCGCACACTTAACATTTTCTACACCGGCACCTAAA 2657
Db	361 TTCGGTGGAGCGGTATTTACTACGCGCACACTTAACATTTTCTACACCGGCACCTAAA 420
Qy	2658 AATTGACGGAAAGCGCGCGCCACCCAAAAACCTTGTGAAAGTGAGGACCCAACTGGGCT 2717
Db	421 AATTGACGGCAAGCGCGCGCCACCCAAAAACCTTGTGAAAGTGAGGACCCAACTGGGCT 480
Qy	2718 GATGGCGGCATTCATCCCGTTCGCTTAAAAATTCGCTTATCGACGACCCGCCACGCG 2777
Db	481 GATGGCGGCATTCATCCCGTTCGCTTAAAAATTCGCTTATCGACGACCCGCCACGCG 540

```
QY 2778 TTTCACACCCATTACCGGATCCCATGATCAGCCCTGATGGTGTGAACATGCT 2837
Db 541 TTTCACACCCATTACCGGATCCCATGATCAGCCCTGATGGTGTGAACATGCT 600
QY 2838 TCTTGGGGCCCAACGCGAAACCTCACCAGTGTGAGCGGTTCTATACCGCTCGACAGATCT 2897
Db 601 TCTTGGGGCCCAACGCGAAACCTCACCAGTGTGAGCGGTTCTATACCGCTCGACAGATCT 660
QY 2898 TGAAGACTGGGAATCTCCGGTGAATCACCCTTGCAGCTCAGTGATGACACACCTGGTTC 2957
Db 661 TGAAGACTGGGAATCTCCGGTGAATCACCCTTGCAGCTCAGTGATGACACACCTGGTTC 720
QY 2958 TGTCTGTGATCTCTCCGATGGCTACATGTGGGAATGCCCAACCTTTTACGCTTCG 3017
Db 721 TGTCTGTGATCTCTCCGATGGCTACATGTGGGAATGCCCAACCTTTTACGCTTCG 780
QY 3018 CGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTTCTGCCAAGGATTTGACCG 3077
Db 781 CGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTTCTGCCAAGGATTTGACCG 840
QY 3078 AATCCAGCATGAGGTTACTCAGTACGCAAGCTCTGACAGTGGGATATGCTGTCGACAA 3137
Db 841 AATCCAGCATGAGGTTACTCAGTACGCAAGCTCTGACAGTGGGATATGCTGTCGCA 900
QY 3138 GCTTGAAGAACTGGCGTCTGCGAGATTCAGGAGATTCAGGAGCTGGATTTGCGCATGA 3197
Db 901 GCTTGAAGAACTGGCGTCTGCGAGATTCAGGAGATTCAGGAGCTGGATTTGCGCATGA 960
QY 3198 ATTCTAGCACCCAGGTTGCACTAAACGTTCTGATGCTGCTGCTGGTGGTGGATGG 3257
Db 961 ATTCTAGCACCCAGGTTGCACTAAACGTTCTGATGCTGCTGCTGGTGGTGGATGG 1020
QY 3258 GCTGCCGCGCAGATGATCACCACAGTGTGACAGGAGGATGGTGCAGTGCCTGAC 3317
Db 1021 GCTGCCGCGCAGATGATCACCACAGTGTGACAGGAGGATGGTGCAGTGCCTGAC 1080
QY 3318 TGTGCCCGCAAGCTTCAATTTGGCGCAACACGAGTCTACCAAGAGCTCTTCTCCGACA 3377
Db 1081 TGTGCCCGCAAGCTTCAATTTGGCGCAACACGAGTCTATCAAGAGCTCTTCTCCGACA 1140
QY 3378 GGGGAGTCGGGGTAAATCAGATCTGTATAGTCTTGAACCTGTCCGAGTAGACATCG 3437
Db 1141 GGGGAGTCGGGGTAAATCAGATCTGTATAGTCTTGAACCTGTCCGAGTAGACATCG 1200
QY 3438 AGCAATATTTCCCTCGAGTGGATGGTGTCCGTTTGTGTGATCGTGTGATGATCG 3497
Db 1201 AGCAATATTTCCCTCGAGTGGATGGTGTCCGTTTGTGTGATCGTGTGATGATCG 1260
QY 3498 TCGGAGTACGAGTAAACCTGGCAATTAATAGTATCGGAGGAGTAATACAGCAATGA 3557
Db 1261 TCGGAGTACGAGTAAACCTGGCAATTAATAGTATCGGAGGAGTAATACAGCAATGA 1320
QY 3558 GATAACTGCAGGTGATGGACAG 3579
Db 1321 GATAACTGCAGGTGATGGACAG 1342
```

## RESULT 9

AAH67870

ID AAH67870 standard; DNA; 1299 BP.

XX AC

XX AC

XX AC

XX AC

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 2905.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PN

XX

PD

XX

PF

XX

XX

PR

PR

PR

PR

XX

PA

XX

PI

PI

XX

XX

DR

DR

XX

XX

PT

PT

PT

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity

Matches 1286;

Conservative

0; Mismatches

12; Indels

1; Gaps

1;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW ) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI: 2001-376931/40.

P-PSDB; AAG92651.

Novel polynucleotides derived from Coryneform bacteria, for identifying

mutation point of a gene, measuring expression of a gene, analysing

expression profile or pattern of a gene and identifying homologous gene

Claim 8; SEQ ID NO: 2905; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein

sequences from the Coryneform bacterium Corynebacterium glutamicum. These

are useful for identifying the mutation point of a gene derived from a

mutant of coryneform bacterium, measuring expression amount and

analysing the expression profile or expression pattern of a gene derived

from Coryneform bacterium, and identifying a homologue of a gene derived

from coryneform bacterium. Coryneform bacteria are useful for producing

amino acids, nucleic acids, vitamins, saccharides and organic acids,

particularly L-lysine. The present sequence is a nucleic acid described

in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from the

European Patent Office.

Sequence 1299 BP; 287 A; 363 C; 350 G; 299 T; 0 other;

Query Match

21.2%; Score 1267.8; DB 22; Length 1299;

Best Local Similarity

99.0%; Pred. No. 0;

Matches 1286; Conservative

0; Mismatches

12; Indels

1; Gaps

1;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db



CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. App, NADPH).

XX  
 SQ Sequence 1287 BP; 325 A; 349 C; 326 G; 287 T; 0 Other;

Query Match 20.7%; Score 1237.4; DB 22; Length 1287;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 242 AACAGAGCCCTCAAGTCGGAAGATAATTAACTAGATCCGTAGACATAAGACATCATAC 301  
 DB 1 AACAGAGCCCTCAAGTCGGAAGATAATTAACTAGATCCGTAGACATAAGACATCATAC 60  
 QY 302 GTCCTATGCTTCTGGAAGAACCAATAAATCTCAGAAAGATGCGAGAAGTGGTGCAATTA 361  
 DB 61 GTCCTATGCTTCTGGAAGAACCAATAAATCTCAGAAAGATGCGAGAAGTGGTGCAATTA 120  
 QY 362 TCAAGAAATGTCAGGTCAAGCAGTTTAAAAAATTCAGGGAAGAAATGTTCCCCCTCGG 421  
 DB 121 TCAAGAAATGTCAGGTCAAGCAGTTTAAAAAATTCAGGGAAGAAATGTTCCCCCTCGG 180  
 QY 422 GGTGATGATGCTTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGGAGAACC 481  
 DB 181 GGTGATGATGCTTTCTCCAACTCGAAACGGCATCATCAGGAACTCTCTGGAGAACC 240  
 QY 482 AGCACCTAAAAACGAGGATTCACCCGGAATCCACCCGAACTCCCGAGATTGTTCCGGTTTATTGA 541  
 DB 241 AGCACCTAAAAACGAGGATTCACCCGGAATTCACCCGGAATTCACCCGAGATTGTTTATTGA 300  
 QY 542 TTTTCAATACGGTGGAAACGGTGGCGGTTTCTACGGGAACGACGAGCCAGGCGAG 601  
 DB 301 TTTTCAATACGGTGGAAACGGTGGCGGTTTCTACGGGAACGACGAGCCAGGCGAG 360  
 QY 602 GAACCGCGCAGTATCACCGGAACATGCGCAGACCTGATGTTGCCAAGCATGGTTTC 661  
 DB 361 GAATCGCGCAGTATCACCGGAACATGCGCAGACCTGATGTTGCCAAGCATGGTTTC 420  
 QY 662 GCGCGCGGTGACGACTGGCAGCGGAGTGGAAACCTTATTCCTTGTGTGAAGAGT 721  
 DB 421 GCGCGCGGTGACGACTGGCAGCGGAGTGGAAACCTTATTCCTTGTGTGAAGAGG 480  
 QY 722 CTGCTGTGCGCATTCACCTCGAGGCGCTTTCATCAGCGATCCGCTGTGGTGCTCA 781  
 DB 481 CTGCTGTGCGCATTCACCTCGAGGCGCTTTCATCAGCGATCCGCTGTGGTGCTCA 540  
 QY 782 AAACCGGATTTATTTTCCCGGCAACCCACAGATCTTGCCCGGCTGATCCATGCGGG 841  
 DB 541 AAACCGGATTTATTTTCCCGGCAACCCACAGATCTTGCCCGGCTGATCCATGCGGG 600  
 QY 842 AANAGTTGGATCAATTCATCAGTACGCGCGGGAACATGACATCTTCTGAGCTTCT 901  
 DB 601 AANAGTTGGATCAATTCATCAGTACGCGCGGGAACATGACATCTTCTGAGCTTCT 660  
 QY 902 CGATCTCTGCGAGCGCACCATCATCTTCTTCCCTTGGGCACTGATGACAGATTTGA 961  
 DB 661 CGATCTCTGCGAGCGCACCATCATCTTCTTCCCTTGGGCACTGATGACAGATTTGA 720  
 QY 962 TACCACTACAGGCAATTCCTTGGCTTAAAGAGAAATGTGAGGGTACCGGTACGCA 1021  
 DB 721 TACCACTACAGGCAATTCCTTGGCTTAAAGAGAAATGTGAGGGTACCGGTACGCA 780  
 QY 1022 TTTGTTCAATGGATGCTCCGCTGCATCATAGGCTCCCGGCGGCGCTTGGCT 1081  
 DB 781 TTTGTTCAATGGATGCTCCGCTGCATCATAGGCTCCCGGCGGCGCTTGGCT 840  
 QY 1082 TGTGCGGCACTGCGGGGACGCATPATGTTGATGTCGCGGCGGCTGCAATTTGGC 1141  
 DB 841 TGTGCGGCACTGCGGGGACGCATPATGTTGATGTCGCGGCGGCTGCAATTTGGC 900  
 QY 1142 CGATGGAACGGTGCATCTAGCTCGTTCCTCAACAGCGCTTTTTCATCAGCGGCGCATGA 1201  
 DB 1201 CGATGGAACGGTGCATCTAGCTCGTTCCTCAACAGCGCTTTTTCATCAGCGGCGCATGA

DB 901 CGATGGAACGGTGCATCTAGCTCGTTCACAAACAGCGCTTTTTCATCAGCGACCCATGGA 960  
 QY 1202 AGCGCCGCGAATGCGACAGCGTGAATACATTTTGGCGTITTTGAAGCTCACCGTACCCGA 1261  
 DB 961 AGCGCCGCGAATGCGACAGCGTGAATACATTTTGGCGTITTTGAAGCTCACCGTACCCGA 1020  
 QY 1262 TGAAGTCCGCGCTTGGCGGATGCGGCGCATCGCGGGGGACACGACATAGCGAG 1321  
 DB 1021 TGGCGTCCGCGCTTGGCGGATGCGGCGCATCGCGGGGGTACCAGCACATAGCGAG 1080  
 QY 1322 TCAGTTCTGTCACACACGTCGCGAGGGTATGACGCTTATCGAGCGACCCCTCCACACCTC 1381  
 DB 1081 TCAGTTCTGTCACACACGTCGCGAGGGTATGACGCTTATCGAGCGACCCCTCCACACCTC 1140  
 QY 1382 AACCGTCCGCGTAAATTTCTCGTCTGGCGATGACGAAATCGTAAATCCAAACCGTGC 1441  
 DB 1441 AACCGTCCGCGTAAATTTCTCGACTTAGCGATACGAAATCGTAAATCCAAACCGTGT 1200  
 QY 1442 AAATTTTGTGCTCTTTGACTCAACAGCGCAGGTGCAAAAGGTCCATTTAGGTTCATCAAGT 1501  
 DB 1201 AAATTTTGTGCTCTTTGACTCAACAGCGCAGGTGCAAAAGGTCCATTTAGGTTCATCAAGT 1260  
 QY 1502 ACTTTAAGTACGAGTAAACTATCCTG 1528  
 DB 1261 AATTTAAATACGAGCAAAACTTTCTCTG 1287

RESULT 11  
 RAAF71567  
 ID AAF71567 standard; DNA; 1287 BP.  
 XX  
 AC AAF71567;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:415.  
 XX  
 KW Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100844-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00943.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031412.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031431.  
 PR 08-JUL-1999; 99DE-1031433.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031562.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 09-JUL-1999; 99DE-1032180.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 09-JUL-1999; 99US-0143208.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 14-JUL-1999; 99DE-1033005.

PR 27-AUG-1999; 99DE-1040765.  
 PR 31-AUG-1999; 99US-0151572.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.  
 XX (BADI ) BASF AG.  
 XX  
 XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;  
 XX WPI; 2001-061975/07.  
 DR P-PSDB; AAB79450.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -  
 XX  
 PS Claim 3; Page 739-740; 1246pp; English.  
 XX  
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (ii) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins  
 CC (iii) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diptheriae in a subject. (i), (ii), (iii) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).  
 XX  
 SQ Sequence 1287 BP; 325 A; 349 C; 326 G; 287 T; 0 other;  
 Query Match 20.7%; Score 1237.4; DB 22; Length 1287;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
 QY 242 AACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAAGACATCATAC 301  
 DB 1 AACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAAGACATCATAC 60  
 QY 302 GTCCTATGCTTCTGGAAGAACCAATTAACCTCAGAAAGATGCGAAGTGTGTCATTA 361  
 DB 61 GTCCTATGCTTCTGGAAGAACCAATTAACCTCAGAAAGATGCGAAGTGTGTCATTA 120  
 QY 362 TCAGAAAATCCAGGTCAAGAGTTAAAAAATTTGAGGAGAAATTTGTCCTCCCTCCG 421  
 DB 121 TCAGAAAATCCAGGTCAAGAGTTAAAAAATTTGAGGAGAAATTTGTCCTCCCTCCG 180  
 QY 422 GGTGATGATGGCTTTCTCAACCTCGAAACGGGATCATCACGGAATCTCTGGAGAAC 481  
 DB 181 GGTGATGATGGCTTTCTCAACCTCGAAACGGGATCATCACGGAATCTCTGGAGAAC 240  
 QY 482 AGCACCTTAAACCGAGATTCCACCCGAACTCCCGCATGTTTCCCGTTTATTGA 541  
 DB 241 AGCACCTTAAACCGAGATTCCACCCGAACTCCCGCATGTTTCCCGTTTATTGA 300  
 QY 542 TCTTCATATCACCGTGGAAACGGTGGCGGTTTCTTACGGGACGACGACGAGGAG 601  
 DB 301 TCTTCATATCACCGTGGAAACGGTGGCGGTTTCTTACGGGACGACGAGGAGGAG 360

QY 602 GAACACCGCGCAGTATACCCGCGAACAATGGCAGCACCGGTGATGTTGCCAAGCATGGTTTC 661  
 DB 361 GAATCGCGCGCAGTATACCCGCGAACAATGGCAGCACCGGTGATGTTGCCAAGCATGGTTTC 420  
 QY 662 GCGCGCGCTGACGCACTGGCAGCGCAGGTGGAACCTTATCCCTTGTGTGAAGAGT 721  
 DB 421 GCGCGCGCTGACGCACTGGCAGCGCAGGTGGAACCTTATCCCTTGTGTGAAGAGG 480  
 QY 722 CTTGCTGTGCGGCATTACCTCGAGGGCCCTTTATCAACGCATCCGCTTGTGTGTCGA 781  
 DB 481 CTTGCTGTGCGGCATTACCTCGAGGGTCTTTTATCAACGCATCCGCTTGTGTGTCGA 540  
 QY 782 AAACCCGATTTTCATTTTCCCGGCAACCAACAGATCTTCCCGGGGTGATCCATCGCGG 841  
 DB 541 AAACCCGATTTTATTTTCCCGGCAACCAACAGATCTTCCCGGGGTGATCCATCGCGG 600  
 QY 842 AAAAGTTTGGATCAATCGATCAGCTAGCGCGGAAACTGACAACTCTTTCTGACCTTCT 901  
 DB 601 AAAAGTTTGGATCAATCGATCAGCTAGCGCGGAAACTGACAACTCTTTCTGACCTTCT 660  
 QY 902 CGATCTCTCGCAGCGCACCATCATGCTTCTTCCCTCGGGCACACTGATGACAGATTTGA 961  
 DB 661 CGATCTCTCGCAGCGCACCATCATGCTTCTTCCCTCGGGCACACTGATGACAGATTTGA 720  
 QY 962 TACCACCTACAGCGCAATTTCCCTTGGCTAAAGAGAAAATGTGACGGTCACGGCTACGCA 1021  
 DB 721 TACCACCTACAGCGCAATTTCCCTTGGCTAAAGAGAAAATGTGACGGTCACGGCTACGCA 780  
 QY 1022 TTTGTTCAATGGATGCTTCCGCTCGCTGATAGGCGTCCCGCAGCGTGGGCGCTTGGT 1081  
 DB 781 TTTGTTCAATGGATGCTTCCGCTCGCTGATAGGCGTCCCGCAGCGTGGGCGCTTGGT 840  
 QY 1082 TGTGTCGGCACGTGCGCGGAGCGCATATGTTGAGTTGATCGCGCAGCGGTGCATTTGGC 1141  
 DB 841 TGTGTCGGCACGTGCGCGGAGCGCATATGTTGAGTTGATCGCGCAGCGGTGCATTTGGC 900  
 QY 1142 CGATGAAACGGTGCATGCTAGCTGTTTCCAAACAGCGCTTTTTCATCAGGACGCGCATGGA 1201  
 DB 901 CGATGAAACGGTGCATGCTAGCTGTTTCCAAACAGCGCTTTTTCATCAGGACGCGCATGGA 960  
 QY 1202 AGCGCGCGGAATGCGCAGAGCGGTGAGTACATTTTGGCGTTTGAACGTCAACGTCACCGGA 1261  
 DB 961 AGCGCGCGGAATGCGCAGAGCGGTGAGTACATTTTGGCGTTTGAACGTCAACGTCACCGGA 1020  
 QY 1262 TGGAGTTCGCCCTGTCGCGGATGGCGGCGCCATCGCGGGGGCACACGACACTAGCGAG 1321  
 DB 1021 TGGCGTTCGCCCTGTCGCGGATGGCGGCGCCATCGCGGGGGTACGACGACACTAGCGAG 1080  
 QY 1322 TCAGTTTCGTGCACGCTGCGCAGGGGTATGACGTTATCGACGCGACCCCTCCACACCTC 1381  
 DB 1081 TCAGTTTCGTGCACGCTGCGCAGGGGTATGACGTTATCGACGCGACCCCTCCACACCTC 1140  
 QY 1382 AACCGTCGCGCTAAATTTCTCGGTCTTGGCGATCAGAAATCGGTAATCAACACCCCTGC 1441  
 DB 1141 AACCGTCGCGCTAAATTTCTCGGACTTAGCGATCAGAAATCGTAAATCAACACCCCTGT 1200  
 QY 1442 AAATTTGCTGCTTTGACTCAACGCGCAGGTGGAAGGTCCTATTAGGTCATCAAGT 1501  
 DB 1201 AAATTTGCTGCTTTGACTCAACGCGCAGGTGGAAGGTCCTATTAGGTCATCAAGT 1260  
 QY 1502 ACTTTTAAGTACGAGTAAATCTATCTCTG 1528  
 DB 1261 AATTTAATACGAGCAAACTTTCTCTG 1287

RESULT 12  
 AAH67872  
 ID AAH67872 standard; DNA; 1152 BP.  
 XX  
 AC AAH67872;  
 XX  
 DT 26-SEP-2001 (first entry)

XX	C glutamincium coding sequence fragment SEQ ID NO: 2907.
DE	
XX	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW	organic acid synthesis; ds.
XX	
OS	Corynebacterium glutamicum.
PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PP	18-DEC-2000; 2000EP-0127688.
XX	
PF	16-DEC-1999; 99JP-0377484.
PR	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
XX	
XX	(KYOW ) KYOWA HAKKO KOGYO KK.
PA	
XX	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI	WPI: 2001-376931/40.
DR	P-PSDB; AAG92653.
DR	
PT	Noval polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	.
XX	
PS	Claim 8; SEQ ID NO: 2907; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a nucleic acid described
CC	in the exemplification of the invention.
CC	Note: the sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SQ	Sequence 1152 BP; 273 A; 320 C; 299 G; 260 T; 0 other;
	Query Match 18.6%; Score 1110.4; DB 22; Length 1152;
	Best Local Similarity 97.7%; Pred. No. 0;
	Matches 1126; Conservative 0; Mismatches 26; Indels 0; Gaps 0
QY	354 GPGCATTATCAAGAAATGCGGTCAAGCAGTTTAAAAAATTGAGGAAGAATTTGCC 413
Db	
	1 GPGCATTATCAAGAAATGCGGTCAAGCAGTTTAAAAAATTGAGGAAGAATTTGCC 60
QY	414 CCCCTCGGGTGATTGATGGTTTTCTCCAACTCGAAAACGGCATATCACGAACCTCT 473
Db	
	61 CCCACGGGGTGATTGATGGTTTTCTCCAACTCGAAAACGGCATATCACGAACCTCT 120
QY	474 GGAGAACGAGCACTTAAAGCGAGGATTCACCCGGAATCCCCACGATTTGCCCGT 533
Db	
	121 GGAGAACGAGCACTTAAAGCGAGGATTCACCCGGAATCCCCACGATTTGCCCGT 180
QY	534 TTATTATGATCTCATATATACGCGTGGAAACGGTGGCGGTTTTCTACGGGAACGAGGAC 593
Db	
	181 TTATTATGATCTCATATATACGCGTGGAAACGGTGGCGGTTTTCTACGGGAACGAGGAC 240
QY	594 CAGGCGGAGAACACCGGCAGTATATCAACGGGAACATGGCAGCACCGTGTATGTTGCCAAGC 653
Db	
	241 CAGGCGGAGGAATGCGCGCAGTATATCAACGGGAACATGGCAGCACCGTGTATGTTGCCAAGC 300

OS Corynebacterium glutamicum.  
 PN WO200102583-A2.  
 XX  
 XX  
 PD  
 XX  
 PF  
 XX  
 PF 11-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-IB00973.  
 XX  
 PR 01-JUL-1999; 99US-0142691.  
 PR 23-AUG-1999; 99US-0150310.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042097.  
 XX  
 XX (BADI ) BASF AG.  
 XX  
 XX Pompeius M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 PI WPI; 2001-080989/09.  
 XX  
 DR  
 XX  
 XX  
 PT Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;  
 PT sugar phosphotransferase system proteins or their portions, useful for  
 PT typing or identifying C. glutamicum or related bacteria, and as markers  
 PT for transformation -  
 XX  
 XX Claim 5; Page 103-104; 144pp; English.  
 PS  
 XX The present invention relates to Corynebacterium glutamicum  
 CC phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.  
 CC The PTS nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for mediating fine chemical production in C. glutamicum or  
 CC related bacteria, the typing or identification of C. glutamicum genome,  
 CC related bacteria, as reference points for mapping C. glutamicum genome,  
 CC and as markers for transformation.  
 XX  
 SQ Sequence 1109 BP; 236 A; 278 C; 303 G; 292 T; 0 other;  
 Query Match 17.6%; Score 1049.8; DB 22; Length 1109;  
 Best Local Similarity 96.7%; Pred. No. 3.3e-306;  
 Matches 1072; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
 QY 4676 TATGATTCGGTGTCCAGTCGGGGTGTGCTTTTCGGTCTGCTACTCACCAGTTCGTT 4735  
 DB 1 TATGATTCGGGGTCCAGTCGGGGTGTGCTTTTCGGTCTGCTACTCACCAGTTCGTT 60  
 QY 4736 ATCACTGTCTGCACCACTCTCCGCCAATTCAGCTGGAGCTTCAACACCGGTGGA 4795  
 DB 61 ATCACTGTCTGCACCACTCTCCGCCAATTCAGCTGGAGCTTCAACACCGGTGGA 120  
 QY 4796 TCCCTTCATCTCGCAACCGCATCCATGGCCCAATATCGCCAGGGTGGCATGTTGGCA 4855  
 DB 121 TCCCTTCATCTCGCAACCGCATATGSGCTAATATCGCCAGGGTGGCATGTTGGCA 180  
 QY 4856 GTGTTCTTCCTAGCGAAGAGTGAAGCTCAAGGCCCTTCAGGTGCTTCAGGTGCTTCC 4915  
 DB 181 GTGTTCTTCCTAGCGAAGAGTGAAGCTCAAGGCCCTTCAGGTGCTTCAGGTGCTTCC 240  
 QY 4916 GCTGTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGGCTCGCTGGCGG 4975  
 DB 241 GCTGTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGGCTCGCTGGCGG 300  
 QY 4976 TTCTACATGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTCACCTTTTGAATC 5035  
 DB 301 TTCTACATGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTCACCTTTTGAATC 360  
 QY 5036 AAGGCAGTTCGGTGGCGCTGCAGGTTCTTCGGTGTGTTCTTATGATGCTCCAGAT 5095  
 DB 361 AAGGCAGTTCGGTGGCGCTGCAGGTTCTTCGGTGTGTTCTTATGATGCTCCAGAT 420  
 QY 5096 ATGGTCATGTTCTTGGTGGCGGTAGTTACCTTTGTATCGCATTCGGCGCAGCGATT 5155  
 DB 421 ATGGTCATGTTCTTGGTGGCGGTAGTTACCTTTGTATCGCATTCGGCGCAGCGATT 480

QY 5156 GCTTATGGCTTTACTTGGTTCGGCGACGACGACGATTCAGATTCAGACGCTGCT 5215  
 DB 481 GCTTATGGCTTTACTTGGTTCGGCGACGACGACGATTCAGATTCAGACGCTGCT 540  
 QY 5216 CCAGTGCCTGCAGGAACGACCAAGCCGAGAGCAGACCCGCGAGATTTTCAACAGAT 5275  
 DB 541 CCAGTGCCTGCAGGAACGACCAAGCCGAGAGCAGACCCGCGAGATTTTCAACAGAT 600  
 QY 5276 TCCACCATCATCCAGGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCTCAGCGAT 5335  
 DB 601 TCCACCATCATCCAGGCACCTTTGACCGGTGAAGCTATTCGCACTGAGCAGCTCAGCGAT 560  
 QY 5336 GGCATGTTTGGCCAGCGGAAAGCTTGGCTCAGGTGTTGGCATCGTCCCAACCAAGGGGAG 5395  
 DB 661 GGCATGTTTGGCCAGCGGAAAGCTTGGCTCAGGTGTTGGCATCGTCCCAACCAAGGGGAG 720  
 QY 5396 CTGGTTTCCACAGTGAGCGGAAAGATCGGTGGTGGCTTCCCATCTGCTCAGCTTTCGCA 5455  
 DB 721 TTAGTTTCTCCGGTGAAGATTTGGTGGCATTTCCCATCTGCGCATGCTTTTCGCA 780  
 QY 5456 GTCCGCACTAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGAC 5515  
 DB 781 GTTCGCACCAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGAC 840  
 QY 5516 ACCGTTAAACCTCAACGGCAGCGACTTTAAACCGCTGAGAGACGAGGCGATGAAGTCAA 5575  
 DB 841 ACAGTAAACCTCAACGGCAGCGACTTTAAACCGCTGAGAGACGAGGCGATGAAGTCAA 900  
 QY 5576 GCAGGGGAGCTGCTGTGAATTCGATATTGATGCAATTAAGCTTCGAGGTTATGAGGTA 5635  
 DB 901 GCAGGGGAGCTGCTGTGAATTCGATATTGATGCAATTAAGCTTCGAGGTTATGAGGTA 960  
 QY 5636 ACCACGCCGATGTTGTTTCGAATACAGAGAAACCGGACCTGTAACACTTACGGTTTG 5695  
 DB 961 ACCACGCCGATGTTGTTTCGAATACAGAGAAACCGGACCTGTAACACTTACGGTTTG 1020  
 QY 5696 GGCAGAAATGAAGCGGAGCCAAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCAGCA 5755  
 DB 1021 GGCAGAAATGAAGCGGAGCCAAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCAGCA 1080  
 QY 5756 ACACCATAGTTGAAAACCTTGAGTGTTCG 5784  
 DB 1081 ACACCATAGTTGAAAACCTTGAGTGTTCG 1109  
 RESULT 14  
 AAF71571  
 ID AAF71571 standard; DNA; 882 BP.  
 XX AC AAF71571;  
 XX AC AAF71571;  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:423.  
 XX  
 KW Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 PN WO200100844-A2.  
 XX  
 XX 04-JAN-2001.  
 XX 23-JUN-2000; 2000WO-IB00943.  
 XX 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031412.

```

PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI ) BASF AG.
XX
PI Pompe'us M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
XX P-PSDB; AAB79454.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 3; Page 751-752; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (ii) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins
CC (iii) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diptheriae in a subject. (i), (ii), (iii) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 882 BP; 259 A; 232 C; 213 G; 178 T; 0 other;
Query Match 12.8%; Score 762; DB 22; Length 882;
Best Local Similarity 93.0%; Pred. No. 3.6e-219;
Matches 820; Conservative 0; Mismatches 50; Indels 2; Gaps 2;
QY 1449 GTGCTCTTTGACTCAAGCGCCAGGTGCAAAAGTCCATTAGTGTCAAGTACTTTAA 1508
DB 1 GTGCTCTTTGACTCAAGCGCCAGGTGCAAAAGTCCATTAGTGTCAAGTACTTTAA 60
QY 1509 GTACGAGTAAACTATCTGTAT-TTAAAGGAGTCCCAACCATGGAATCACTATCTGCA 1567

```

```

Db 61 ATACGAGCAAAACTTTCTGTGATAATAAAGGAGTCCGACCATGGACATCATCATCTGCAA 120
QY 1568 AGACGAGCAAGAAGTCGCAAAAGCAGTTGTCAGTCTTAATCGCACCTTTGCGCAAGAGG 1627
Db 121 AGACGAGCAAGAAGTCGCAAAAGCAGCGCCCTGATCGCACCTTCGCAACATGAGG 180
QY 1628 TGGAACTTTGGGGCTTGCACAGGATCTCACCATTGAGTACCTACCAAGAGCTCATTCG 1687
Db 181 CGGAACCTTTGGGGCTTGCACAGTGGATCTCACTTTGAGCACCTACCAAGAGCTCATTCG 240
QY 1688 CATGTATGAAGCTGGGGAAGTGTCAAGAACTGCAAGGCAATCTTTGTTGGATGAATA 1747
Db 241 CATGTATGAAGCTGGGGAAGTGTCAAGAACTGCAAGGCAATCTTTGTTGGATGAATA 300
QY 1748 COTGGGACTAACCCGTGAGGATGAACAGCTACTTTTAAACCATTCGCAAGAGATTAC 1807
Db 301 CGTGGATTAAACGCGCAGCATGAAACAGCTACTTTCAAACCATTCGTAAGAGATTAC 360
QY 1808 TGACCACATCGACATCGTTGATGAAGAGTCTTACAGCCAGATGTTGCAAAACCTTGATCC 1867
Db 361 TGACCACATCGACATCGTTGATGAAGAGTCTTACAGCCAGATGTTGCAAAACCTTGATCC 420
QY 1868 ATACGAAGCAGCTGCAGAGTATGAGGCAAGATGCTCTGCAAGTCCGTTGAAGTTCAAT 1927
Db 421 ATACGAAGCAGCTGCAGAGTATGAGGCAAGATGCTCTGCAAGTCCGTTGAAGTTCAAT 480
QY 1928 CCTTGGCATCGCGGAAACGG-CACATGCTTTTATTGAACCATCATCTTCTCTGTCAAG 1986
Db 481 CCTTGGCATCGCGGAAACGGCCACATCGCTTTCAATGAGCCATCATCTTCTCTGTCAAG 540
QY 1987 ACTGACAAAGTCCAGGCGCTGCACCTTAAACTGTGGAGGACAAACGCTCGATTTCAA 2046
Db 541 ACTGACAAAGTCCAGGCGCTGCACCTTAAACTGTGGAGGACAAACGCTCGATTTCAA 600
QY 2047 CACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGCACTTTGTCGCGCGC-2106
Db 601 CACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGCACTTTGTCGCGCGC 660
QY 2107 GCAAAACATCGTGTGTTGGTGGCAACTGTTGAAGGAAAGCCGACGCACTCCGCGGAAC7GT 2166
Db 661 GCAAAACATCGTGTGTTGGTGGCAACTGTTGAAGGAAAGCCGACGCACTCCGCGGAAC7GT 720
QY 2167 GGAAGCCCAAGTACTGCTTCTTCCCGAGTTCATCTGTAGATGCACATGCCACC 2226
Db 721 GGAAGCCCAAGTACTGCTTCCCGAGTTCATCTGTAGATGCACATGCCACC 780
QY 2227 ATCATCGTTGGATGAAGCAGCAGTATCCCAAGCTGAAACCGCTGATCACTACCGTCTCAT 2286
Db 781 CATCATCGTTGATGAAGCAGCAGTATCCCAAGCTGAAACCGCTGATCACTACCGTCTCAT 840
QY 2287 GGAGCAATTAAGTCCGCTAGAAACAAAGGAAAGTACTG 2328
Db 841 GGAGCAATTAAGTCCGCTAGAAACAAAGGAAAGTACTG 882

```

## RESULT 15

AAH67871

ID AAH67871 standard; DNA; 759 BP.

XX AAH67871;

XX AAH67871;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 2906.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

PN

XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 18:38:12 ; Search time 7503 Seconds

(without alignments)  
19335.388 Million cell updates/sec

Title: US-10-019-284A-1

Perfect score: 5969

Sequence: 1 agtcgtgcagccaccatt.....tgatatatttgcgcgcgtgaa 5969

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pla:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	174.8	2.9	917	29	BZ685850 PUBES66TD
C 2	123.6	2.1	890	14	CD499281 CDA38-D12
3	123.6	2.1	1030	14	CD499280 CDA38-D12
4	103.8	1.7	818	12	BM017859 603645335

5	103.2	1.7	669	9	AL656223
6	102.6	1.7	881	13	EX458246
7	102.6	1.7	947	13	EX374555
8	102.6	1.7	1051	12	BM476096
9	102.6	1.7	1063	12	BM910583
10	102.6	1.7	1095	9	AL560879
11	102.6	1.7	1201	13	EX406188
12	102.6	1.7	1201	13	EX421863
13	102.4	1.7	1040	29	CNSOINFL
14	102.2	1.7	669	10	BE790435
15	102.2	1.7	847	13	BU171522
16	101.8	1.7	1201	9	AL546345
17	101.4	1.7	1201	13	EX324827
18	101.2	1.7	1201	13	EX428009
19	100.2	1.7	875	13	BU158666
20	100	1.7	746	12	BI758774
21	100	1.7	812	12	BI825295
22	100	1.7	905	13	BU179037
23	100	1.7	934	13	BO929688
24	100	1.7	984	13	BU167870
25	99.6	1.7	712	10	EG431403
26	99.6	1.7	767	10	EG388237
27	99.6	1.7	940	12	BM472534
28	99	1.7	851	14	CB558813
29	99	1.7	901	13	BU904006
30	98.6	1.7	1201	9	AL549852
31	98.2	1.6	776	12	BM016765
32	98.2	1.6	801	13	BU167497
33	98.2	1.6	870	10	BE742200
34	98.2	1.6	1042	12	BM557913
35	97.8	1.6	665	12	BI553285
36	97.8	1.6	757	12	BG771300
37	97.2	1.6	679	12	BG971729
38	96.6	1.6	848	12	BI601355
39	96.6	1.6	942	12	BI760659
40	96.6	1.6	1201	13	EX445414
41	96.6	1.6	775	28	BR398566
42	96.2	1.6	632	9	AL640829
43	95.6	1.6	627	10	BE280307
44	95.6	1.6	647	12	BJ332387
45	95.2	1.6	640	12	BM487491

#### ALIGNMENTS

RESULT 1  
BZ685850/c  
LOCUS PUBES66TD 2M\_0.6\_1.0.KB zea mays genomic clone ZM87A035X11, 917 bp DNA linear GSS 05-FEB-2003  
DEFINITION genomic survey sequence.  
ACCESSION BZ685850  
VERSION BZ685850.1 GI:28245462  
KEYWORDS GSS.  
SOURCE zea mays  
ORGANISM zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 917)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.  
TITLE Maize Genomics Consortium  
JOURNAL Unpublished  
COMMENT Contact: Cathy Whitelaw  
TIGR  
9172 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: Sheared ends.  
Location/Qualifiers

```
source 1. .917
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZM0.6.1.0_KB"
/clone="ZM0.6.1.0_KB"
/notes="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high;
Cor selected genomic DNA library"
BASE COUNT 197 a 287 c 279 g 154 t
ORIGIN

Query Match 2.9%; Score 174.8; DB 29; Length 917;
Best Local Similarity 52.7%; Pred. No. 3e-39;
Matches 483; Conservative 0; Mismatches 412; Indels 21; Gaps 4;

QY 3930 TGAAGGACCTTTGAACTGGGCGCATGTTCAGATCATCGTCGGGCGAGCGATGTGG 3989
Db 914 TCAAGGGTCTGCTTCAACGGGCGCTGTTCAGGTGTAAATGCCCGCGTGAAGTGG 855
QY 3990 ATCATGTTTCAAGAACTCGATGACGACCACTCCAAAGACATCGCTGTTCACAGAGC 4049
Db 854 AGAAGTCTACACCGCTCGGCGGACGACCGGC---CTCGCGCGCGCCACCATCGCGG 798
QY 4050 AGCTCAAAGATGTTGGCTTAACACACCGCAACTGTTTCAGCCGCTGTGCAAGGATTTGG 4109
Db 797 ACGTGAAGAAAGGGCGGCGACAGACCAACGCCCATGAGCGTCTGTGGGTGTTCT 738
QY 4110 CGGACATTTTCGTCGCGCTGATTCCTCAATCTTGTGTTGGCGGTCTGCTCATGGCTATCA 4169
Db 737 CGGAGTGTTCATCGCGATCTCTCCGCGCTGATCATTCGCGGCTGTGTGATGGGCGTGA 678
QY 4170 ACAATGTTGTTGGTGGCGAGATCTGT---TCGGTCCGAATCACATGGTGGAGATTTCC 4226
Db 677 ACAACCTGATGGCGCAAGGCGATGTTTCATCGAGGCGCAGAGCGTCTCGAGCGCTATC 618
QY 4227 CTGATCAGCGGTGTGCTGAGATGATCACTGATGATGCTGCGCGGTCTGCGGTCTT 4286
Db 617 CGAACCTGATGGCTGTGAGGCGCTGATCAACCTGATGGCCACACCTCGTCTGTTCC 558
QY 4287 TGGCAGTGTGTTGTTTTCACCGCAACCAAGCGTTTCGGTGGCAATGATTCCTGGCGG 4346
Db 557 TGGCAGCGGTGTGTTGCTGCGCGCAAGCGCTTGGCGCAGTGAATCTCGGCA 498
QY 4347 CCGCATGTGATGGGATGGTGTTCACACCTGTTACGCGTACGAGCTGGCGCGCA 4406
Db 497 TCGTGTGCTGATGCTGCTGATCCAGACCTGCTCAACGCGCTGGAACCTAGCGCAAG 438
QY 4407 CCATGACCGCGGCGAAATGCCAATGTGTCCTGTTG-----GTTGGATG 4454
Db 437 CGGTGCGGCTCTCGACGGCGCAGAGCTGCGTACTTCGATATTTTCGTTGGTTCAAGA 378
QY 4455 TTGCTAAGCTGTTTACAGGCGACCGTGTCTCTGCTGTTGGTGTCTCTTGGATCTGG 4514
Db 377 TTGAGAAAGTGGCTACAGGGCGAGATCTGCGGATCTGTGATGGCGGCTATGTGATGA 318
QY 4515 CAACGATCGAAGTGTCTCCACAAAGCGATCATGGGCTGCGAGACTTCTCTGATCACCC 4574
Db 317 GCGTATCGAAGTGTCTGCGGCGGAGTGGCCAAACGGATTCATTTGCTGTTGCTG 258
QY 4575 CAGTGTGATCTGCTGCTCACCGGCTTCCTTACGTTCAATGCTATGTTGTCAGCAATGC 4634
Db 257 CGATCACCACTCGTGTACCGGGGTGCTCGCCCTGGCCATTTATCGGCGGTCACCC 198
QY 4635 GCTGGTGGTGTGCTGTCGACACGCTGCTCGAGGACTCTATGATTTGCGGTGTCAG 4694
Db 197 GTCACTTGGCATCTGATCACGAAGCGGTGTGCTACCTTGTTCGACCTGGCACCAGTGG 138
QY 4695 TCGCGGCTGCTGTTTGGTCTGCTTACTACCAATCGTATCACTGTTGCTGACCACT 4754
Db 137 TCGCGCGCGGATTTTGGTCTGCTGTATGCGCGGTGTTGATCATCCGCGCATGCCACA 78
QY 4755 CCTTCCCGCCAAATGAGCTGGAGC---TGTTCAACAGGGTGGATCCTTCACTCTCGCA 4811
```

```
Db 77 TGTTCCTCGCGTGGACTCGAGCTTATTTACTACGGGCGCACCTTCACTGCGCGA 18
QY 4812 CCGCATCCATGGCCAA 4827
Db 17 TGATCGTTATGTCCAA 2

RESULT 2
CD499281/LOCUS
DEFINITION CDA38-D12.Y18-s SHGC-CDA Gasterosteus aculeatus linear EST 12-JUN-2003
ACCESSION CDA38-D12.3', mRNA sequence.
VERSION CD499281.1 GI:31426312
KEYWORDS EST.
ORGANISM Gasterosteus aculeatus (three spined stickleback)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished
Contact: Kingsley, DM
HMMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cgm.stanford.edu
Plate: 38
High quality sequence start: 13
High quality sequence stop: 799.
Location/Qualifiers
1..890
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA38-D12"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoRI
(5' adaptor); Site_2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."
BASE COUNT 194 a 229 c 209 g 257 t 1 others
ORIGIN

Query Match 2.1%; Score 123.6; DB 14; Length 890;
Best Local Similarity 52.3%; Pred. No. 3.7e-24;
Matches 346; Conservative 0; Mismatches 309; Indels 7; Gaps 3;

QY 1606 TCGCACCTTCCCAACAGGCTGACCTGGGCTTGCACAGGATCTCCACACTGA 1665
Db 771 TCCGACCTGGTCCGACAGATTTTACCTGGGCTTCCACAGGAGCACCACATGG 712
```



```

Db      708  CCGCTTCCAGCAGCAGCCGCGACAGACATTTTGTGTGATGAAGAGCGCACATGGGAAC 767
QY      2259  TG 2260
Db      768  TG 769

BM017859      818 bp      mRNA      linear      EST 30-OCT-2001
603645335F1 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5427056 5',
mRNA sequence.
BM017859
EST.
BM017859.1 GI:16532213
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 818)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1892 row: e column: 09
High quality sequence stop: 803.
Location/Qualifiers
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5427056"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_98"
/notes="Organ: brain; Vector: pOTB7; Site1: XhoI; Site2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT      207 a      206 c      221 g      184 t
ORIGIN

Query Match      1.7%; Score 103.8; DB 12; Length 818;
Best Local Similarity 49.5%; Pred. No. 2.5e-18;
Matches 327; Conservative 0; Mismatches 327; Indels 7; Gaps 2;

QY      1610  ACCCTTCGCCAACAGGGTGGACCTTTGGGCTTGGCAACAGGATCCTCACCACCTAGTAC 1669
|||||
Db      78  ACCCAGGCCAGAGAAGTACTTCCCTGGGGCTCCCTCCCTGGGAGTACCCCTGGCTG 137
|||||

QY      1670  CTACCAAGAGCTATTTCGGCATGATCAAGCTGGGGAAGTGTATTCACGAAGCTGCAAGGC 1729
|||||
Db      138  CTACAGAAGCTGATTGAATATACTAAGATGGGACCTGTCTTTAATATATGTGAGAC 197
|||||

QY      1730  ATTCTTGTGGATAGCTGGGAGTAACCCGTGACGATGAAGAAACAGACTACTTTAAAC 1789
|||||
Db      198  CTTCAACATGATGATAGTGGGCTTCTCTCGAGACCACCGGAGAGTTACCACCTCCT 257
|||||

QY      1790  CATTCCCAAGAGTTTCACTGACCATCGACATCGTTGATGAGAGGTCAGAGCCCGCAGA 1849
|||||

```

```

Db      258  CATGTGGAACAACATCTTCTTAAAGCAGCATTTGACATCCACCAGAAACACCCACATTTCTGGA 317
1850  TGGTCCAAACCCCTGATCCATACGAGCAGCTGCAGATATGAGCAGCAAGATCGCTGCAGA 1909
|||||
Db      318  TGGGAATCAGTCTGACCTACAGGAGAAATGTGATGCCCTTTGAAGAGAAATCAAGGCTGC 377
|||||
QY      1910  ATCCGTTGAAGTTCAAAATCCTTTGGCATCGGCGG----AAAGGGCACATCGCTTTTCATTGA 1965
|||||
Db      378  AGGTGGGATCGAGCTATTGTTTGGAGGCATCGGCCCTGATGACACATTTGCTTCAACGA 437
|||||
QY      1966  ACCATCACTTCTCTGTCAGGACTGCACAAGGTCCAGCGCTGCACCCCTAAATCTGTTGGA 2025
|||||
Db      438  GCCAGGCTCCAGTCTGGTGTCCAGAGCCCGTGTGAAGACGCTGGCCATGATACCATCCT 497
|||||
QY      2026  GGACAACGCTCGATTCTTCAA---CACCATCAAGAGGTGCCAACCCACGCGCTCACCCA 2082
|||||
Db      498  GGCCAAATGCTAGGTCTTCGATGGAGAAATCACCAAGGTGCCACCATGGCTTACGGT 557
|||||
QY      2083  GGGTTTGGGCACTTGTCCCGCGCGCAAAACATCGTTGTTGGTGCACTGGTGAAGGAAA 2142
|||||
Db      558  GGGGTGGGCACTGTCTGATGATGCTAGAGAGGTGATGATCCTTATCACAGGTGCTCAAA 617
|||||
QY      2143  ACCCGACCCCATCCGCGGAAGTGTGGAAGGCCCACTGCTCTTCTGCCAGGTTCAT 2202
|||||
Db      618  GGCATTGCTCTGTACAAAGCCATCGAGGAGGAGTGAACACATGTGGACCGTGTCTGC 677
|||||
QY      2203  CCTGTAGATGCACAACATGCCACCATCATCTCTGATGATGAAGCAGCAGATGCCAAGCTGGA 2262
|||||
Db      678  CTTCCAGCAGCATCCCGCACCGGGTTTGTGTGTGACGAGATGCCACCTTGGAGCTGAA 737
|||||
QY      2263  A 2263
Db      738  A 738

RESULT 5
AL656223
LOCUS
DEFINITION      AL656223 XGC-neurula Silurana tropicalis cDNA clone TNeu035007 5',
mRNA sequence.
ACCESSION      AL656223
VERSION        AL656223.1 GI:17668464
KEYWORDS
SOURCE
ORGANISM
Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 669)
Huckle.E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu035007.SP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..669
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/notes="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
EcoRI-NotI cut cDNA was then ligated into pCS107 with

```



cdna clone CSODB007YEL3 5-PRIME, mRNA sequence.

ACCESSION BX374555

VERSION BX374555.1 GI:30456323

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 947)

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3200.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODB007AC07QPlcluster=3200.r. Contact :

Feng Liang Email: fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODB007AC07QPl.

Location/Qualifiers

1..947

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODB007YEL3"

/tissue\_type="NEUROBLASTOMA COT 10-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 255 a 230 c 233 g 209 t

ORIGIN

Query Match 1.7%; Score 102.6; DB 13; Length 947;

Best Local Similarity 49.8%; Pred. No. 6.1e-18;

Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

QY 1632 ACCTTGGGCTTGCACAGGATCCTCACCAGTACTACCAAGAGCTCATTCGCATG 1691

DB 187 ACCCTGGGCTCCCCACATGGGAGTACCCACCTGCTGCTCAAGAAGCTGATTGAATAC 246

QY 1692 TATGAAGCTGGGAAGTGTCTCAATCAAGAACTGCAAGGATCTTGTGGATGAATACGTG 1751

DB 247 TATAAGATGGGAGCTGTCCTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 306

QY 1752 GGACTAACCGTGACGATGAAGAAACAGCTACITTAACACCACTGCGAAGAGTTCACTGAC 1811

DB 307 GGCCCTTCCTCGAGACCCCGGAGAGTACCACCTCTCATGTGGAAACACTTCTTCAAG 366

QY 1812 CACATCGCATCGTTGATGAGAGGTCTACAGCCAGATGTCGCAACCCCTGATCCATAC 1871

DB 367 CACATTCATACCCACAGAAACACCCACATCTCGATGGAAATGCAGTGCACCTACAG 426

QY 1872 GAAGCAGCTGAGATGATGAGCAAGATGCGTGCAGAAATCGTGTGAATCAATCCTT 1931

DB 427 GCAGATGTGTGATGCCTTTTGAAGAGAGATCAAGGCTGCAGGTGGGATGAGTATTTGTT 486

QY 1932 GCATCGGGGG---AAACGGGACATCGCTTTTCATTGAACCATCATCTTCTCTGTCAGGA 1987

DB 487 GGAGCATCGGCCCTGATGGACACATTCCTTCAACGAGCCAGGCTCCAGTCTGGTGTCC 546

QY 1988 CTGACAAAGGTCCAGCGCTGCACCCCTAAACATTTGGAGGACACGCTCGATCTTCTCAA- 2046

DB 547 AGGACCCGTGTGAAGACGCTGGCCATGGATACCATCTCTGGCCAAATGCTAGTCTTCTGAT 606

QY 2047 --CACATCGAGAGGTCCCAACCCACGCGCTCACCCAGGTTTGGGACACTTTGTCGCCG 2104

Db 607 GGAGAACTCACAAGGTGCCCAACCATGGCCTTGACGGTGGGGGTGGGCACCTGTCTATGGAT 666

QY 2105 GCGCAAAACATCGTGTGGTGGCAACTGGTGAAGGAAAGCCGACGCCATCCCGGAAC 2164

Db 567 GCTAGAGAGTGTATCTTATACAGGTGCTCAAGAGCATTTGCTCTGTACAGGCC 726

QY 2165 GTGGAAGCCAGGTGACTGCTTCTGCCAGGTTCATCTCTGTAGATGCACACATGCCA 2224

Db 727 ATCAGGAGGAGTGAACCATGTGACCGTGTCTGCTTCCAGGAGCATCCCGCAC 786

QY 2225 CCATCATCGTGTGATGAAGCAGCAGTATCCAACTGAA 2263

Db 787 GTGTTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 825

RESULT 8

BM476096

LOCUS

DEFINITION AGENCOURT\_6480896 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5580152

5', RNA sequence.

ACCSSION BM476096

VERSION BM476096.1 GI:18525138

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1051)

NIH-MGC http://mgi.nci.nih.gov/.

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12337 row: p column: 09

High quality sequence stop: 648.

Location/Qualifiers

1..1051

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5580152"

/tissue\_type="retinoblastoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_67"

/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Library constructed by Life

Technologies."

BASE COUNT 280 a 253 c 289 g 229 t

ORIGIN

Query Match 1.7%; Score 102.6; DB 12; Length 1051;

Best Local Similarity 49.8%; Pred. No. 6.5e-18;

Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

QY 1632 ACCTTGGGCTTGCACAGGATCCTCACCAGTACTACCAAGAGCTCATTCGCATG 1691

DB 129 ACCCTGGGCTCCCATCTGGAGTACCCACTTGGCTGCTACAGAAGCTGATTGAATAC 188

QY 1692 TATGAAGCTGGGAAGTGTCTATTCAGAACTGCAAGCATTTCTTGTGGATGAATACGTG 1751

DB 189 TATAAGATGGGAGCTGTCTTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 248

QY 1752 GGACTAACCGTGCAGATGAAGACGCTACTTTAAACCATTCGCAAGAGTTCAGTGAC 1811

DB 249 GGCCTTCTCGAGACCAACCCGAGAGTACCACCTCTCTCATGTGGAAACACTTCTTCAAG 308



QY	2225	CCATCATCGTTGGATGAAGCAGCAGTATCCAAGCTGGAA	2263
Db	811	GTGTTTGTCGACGAGGATGCCACCTTGGAGCTGAAA	849
RESULT 11			
EX406188			
LOCUS	EX406188	1201 bp	linear
DEFINITION	BX406188 Homo sapiens FETAL LIVER Homo sapiens CDNA clone		
	CS0DM010YB02 5-PRIME, mRNA sequence.		
ACCESSION	EX406188		
VERSION	EX406188.1	GI:30762669	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jassee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 4334.r For  
 more information about this cluster, see

```

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOAM010DA01QPI&cluster=w4334.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSOAM010DA01QPI.
Location/Qualifiers
source
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM010YB02"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT-6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoSV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 297 a 263 c 298 g 280 t 63 others
ORIGIN

Query Match 1.7%; Score 102.6; DB 13; Length 1201;
Best Local Similarity 49.8%; Pred. No. 7e-18;
Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

Qy 1632 ACCTTGGGCTTGCACAGGATCCTACCACCTGAGTACCTACCAAGAGCTCATTCGCATG 1691
Db 139 ACCCTGGGCTCCCATCTGGAGTACCCCACTTGGCTGCTACAGAGCTGATGAATAC 198

Qy 1692 TATGAAGCTGGGGAAGTGTATTCAGAACTGCAAGGCATCTCTTGTGGGATGAATACGTG 1751
Db 199 TATAAGAATGGGACCTGTCTTTAAATATGTGAAGACCTTCAACATGGGATGAGTACGTG 258

Qy 1752 GGACTAACCGGTGACGATGAAACAGCTACTTTAAACCATTCGCAAGAGTTCACCTGAC 1811
Db 259 GGCCTTCTCGAGACCCCGGAGAGTTACCACTCCTTCATGTGGACAACTTCTTCAAG 318

Qy 1812 CACATGCACATCGTTGATGAAGAGGTCTACAGCCAGATGGTGCAAACCCGTGATCCATAC 1871
Db 319 CACATTGACATTCACCCAGAAAACACCCCACTTCGTGGATGGGAATGCAGTGCACCTACAG 378

Qy 1872 GAAGCAGCTGCAGAGTATGAGGGCAAGATCGCTCGAATCCGTTGAGTTTCAATCCTT 1931

```

```
Db 379 GCAGAAATGATGCCCTTTGAAGAGAGATCAAGGCTGCAGGTGGGATCGAGCTATTGTT 438
QY 1932 GGCATCGGGG---AAACGGACATCGCTTTCATTGAACCATCATCTTCTCTGTCCAGGA 1987
Db 439 GAGGAGATCGGCGCTGATGACACATGCTTCAACGACCGAGCTCCAGTGTGGTCC 498
QY 1988 CTGACAAAGTCTCAGCGCTGCACCCCTAAACTGTGGAGGACAAAGCTCGATCTTCAA- 2046
Db 499 AGGACCGGTGTGAAGACGCTGGCCATGGATACCATCTCTGGCCAAATGCTAGGTCTTCGAT 558
QY 2047 --CACCATCGAAGAGTCCCAACCCACGCGCTCACCGAGGTTGGGCACTTTGTCCCGC 2104
Db 559 GGAGAACTCACCAGGTGCCACCATGGCTTCACGGTGGGGTGGGCACTGTCTGAT 618
QY 2105 GGCACAAACATCGTGTGGTGGCACTGTGGAAGGAAAGCCGACCCCATCCGGGAAGT 2164
Db 619 GCTAGAGAGGTGATGATCTTATACAGAGTGTCAAGGCAATTTGCTCTGTACAGGCC 678
QY 2165 GTGGAAGGCCAGTACTGCTTCTGCCAGGTTTCCATCTGTAGATGACACATGCCA 2224
Db 679 ATCGAGAGGAGTGAACCATGTGGACGCTGTCTGCTTCCAGCAGGATCCCGCCACC 738
QY 2225 CCATCATCGTGTGATGAAGCAGCAGTATCCAAAGCTGGAA 2263
Db 739 GTCTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 777

RESULT 12
BX421863
LOCUS
DEFINITION
BX421863 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
clone CSODH005YE10 5-PRIME, mRNA sequence.
BX421863
BX421863.1 GI:30651259
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODH005BC05QP1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODH005YE10"
/tissue_type="T CELLS (JURKAT CELL LINE)"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
/notes="vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 311 a 267 c 316 g 268 t 39 others
ORIGIN
Query Match 1.7%; Score 102.6; DB 13; Length 1201;
Best Local Similarity 49.8%; Pred. No. 7e-18;
Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;
QY 1632 ACCTTGGGGCTTGCAAGAGATCCTCACCAGTGTAGTACCTACCAAGAGCTCATTCGGATG 1691
```

```
Db 229 ACCCTGGGCTCCCGACTGGGAGTACCCCACTTGCCTGCTACAAGAGCTGATTGAATAC 288
QY 1692 TATGAAGCTGGGAGTGTCTATTCAAGAACTGCAAGGCAATCTTCTTGGATGAATAGTG 1751
Db 289 TATGAAGTGGGACCTGTCTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 348
QY 1752 GGACTAAGCCGTGAGGATGAAGAAAGAGTACTTTAAACCATTTGCAAAAGAGTTCACATGAC 1811
Db 349 GGCCTTCTCGAGACCCCGGAGAGTTACCACTCTTCATGTGGAAACAAGCTTCTTCAAG 408
QY 1812 CACATCGCATCTGTTGATGAAGAGTCTACAGCCGAGATGGTGCACAAACCTTGATCCATAC 1871
Db 409 CACATTGACATCCACCCAGAAACACCCACATTTGGATGGGAATGAGTGCAGTGCACCTACAG 468
QY 1872 GAAGCAGCTGCAGATGATGAGGCAAGATCTGCTGAGAAATCCGTTGAAGTTCATCTT 1931
Db 469 GCAGAATGTGATGCCCTTTTGAAGAGAAGATCAAGGCTCAGGTGGGATCGAGCTATTGTT 528
QY 1932 GGCATCGGGG---AAACGGACATCGCTTTCATTGAACCATCATCTTCTCTGTCCAGGA 1987
Db 529 GGAGCATCGGCGCTGTGAGACACATTCGCTTCAACGAGCAGGCTCCAGTCTGGTGCC 588
QY 1988 CTGACAAAGTCTCAGCGCTGCACCCCTAAACTGTGGAGGACAAAGCTCGATCTTCAA- 2046
Db 589 AGGACCGGTGTGAGACGCTGGCCATGGATACCATCTCTGGCCAAATGCTAGGTCTTTCGAT 648
QY 2047 --CACCATCGAAGAGTCCCAACCCACGCGTCAACCCAGGTTTGGGCACTTTGTCCCGC 2104
Db 649 GGAGAACTCACCAGGTGCCACCATGCTTTCAGGTTGGGGTGGGCACTGTCTGATGAT 708
QY 2105 GCGCAAAACATCGTGTGGTGGCAACTGTGGAAGGAAAGCCGAGCCCATCCGGGAAGT 2164
Db 709 GCTAGAGAGGTGATGATCTTATACAGTGTCTCAAGAGGCAATTTGCTCTGTACAAGGCC 768
QY 2165 GTGGAAGGCCAGTACTGCTTCTTGGCCAGGTTCCATCTGTAGATGACACATGCCA 2224
Db 769 ATCGAGAGGAGTGTGAACCATGTGGACGCTGTCTGCTTCCAGCAGCATCCCGCCACC 828
QY 2225 CCATCATCTGTTGATGAAGCAGCAGTATCCAAAGCTGGAA 2263
Db 829 GTGTTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 867

RESULT 13
CNS01NF1
LOCUS
DEFINITION
CNS01NF1 1040 bp DNA linear GSS 14-JUN-2001
Anopheles gambiae GSS T7 end of clone 23002 of Notre Dame library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
AL152142
AL152142.1 GI:7013061
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Anopheles gambiae (African malaria mosquito)
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 1040)
REFERENCE
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
2 (bases 1 to 1040)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Submitted (16-FEB-2000) BMFI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
```

```
FEATURES
source
Location/Qualifiers
1. .1040
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="23002"
/clone_lib="Notredame1"
/note="end : 17"
BASE COUNT 194 a 314 c 322 g 206 t 4 others
ORIGIN
Query Match 1.7%; Score 102.4; DB 29; Length 1040;
Best Local Similarity 49.0%; Pred. No. 7.4e-18;
Matches 396; Conservative 0; Mismatches 396; Indels 16; Gaps 4;
QY 3819 TTGGCGGGAAGACACATGTGCGCGCGACACTGTGCAACGCGTTTACGCCCTCGGC 3878
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 TGGCGGCGCGGAGAACATCGGCACCGTCAGCCATTGCAATACCGCGCTGCGTTTCGTCC 118
QY 3879 TCAAGACACCAAGGATGTGGATCGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCA 3938
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 TCAATGACCCAGCAGCGCCAGCCCAAGAAATGCAAGAGCTGCCGATGTTCAAGGTT 178
QY 3939 CCTTTCAAACTGGCGCATGTTCCAGATCATCTGGGCGCAGCGATGTGGATCATGTTT 3998
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
179 GCTTACCAACCGCGCGAGTTCAGAGTGGTGTGATCGGCACCGACGTCGCGGACTA----- 233
QY 3999 TCAAGAACTCGATGACGCAACCTCCAAAGACATGCTGTCCACAGCAGCTCAAG 4058
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 -TATCAGCGCTGATCGCCAGCAGCGCGGTCAACGAGCCGACAAAGACAGGCAAGG 292
QY 4059 ATGTTTGGCTAACAGCCCACTGTTTCAGCGCTGCTGTGAAGTATTGGCGGACATTT 4118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 TCGCCGCGCGCCAGAACATGACCTGGACGCGCACCACATCTCTCACTTCGCGGAGATCT 352
QY 4119 TCGTCCGCTGATTCCAATCTTGGTGTGGCGGTCTGCTCATGCTATCAACAATGTGT 4178
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
353 TCTTCCGCTGCTGCCAGCGCTGATCAGCGCGGCTGATCTCTCGCTTCGCAAGTGA 412
QY 4179 TGGTTCGGAGGATCTGTTCGGTCCCAATCACTGTGGAGATGTTCCCTCAGATCAGG 4238
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 TCGCGACATCCCGATGTCGGCGGCAACCGTGGCGGATGCAACCGCGGTGGA 472
QY 4239 GTGTTGTGATGATCAACCTGATGCTATCGCGCTGCGGTTCTTTCAGATGTGG 4298
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 CCATTTACGACTTCCTGTGCTGCTGGTGAAGGATCTCTATGTTCTGCGGTGGCA 532
QY 4299 TTGGTTTCACCGCAACCAAGCGTTTCGGTGGCAATGATTCCTGGCGCGCGGATGTA 4358
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 TCTGTGCTGCGACGGTGAAGATGGGCGGCACCG-GTGTGCGGATCGTCTGCGGCG 591
QY 4359 TGGCGATGTTTCCCAACCTGTTAACGGCTACGACGTGGCGGCCACCATGACGCGG 4418
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 TCACCTGTGTG-TCGCCATGCTCTGATGAATCTTCACTGCTGCGCCAGCAACCGCGT 650
QY 4419 GCGAAATGCGAATGTGGTCCCTGTTGGTTGGATGTTGCTCAAGCTGTTTACCGGCA 4478
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 G-----AAGTGTGGAATCTCGGCTGCTGATCCAGAAAGTCGCGCTATCAGCGCG 702
QY 4479 CCGTGTCTCTGTGCTGTGCTCTTGGATCTTGGCAACGATCGAGAAGTCTCTGCACA 4538
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 AGGTGATCCCGTCTGATCTCGGAGGATGCGCGTGGCGTGGATCGAAACCGCGGTGA 762
QY 4539 AGCGACTCATGGCACTGCAGACTTCTGTATCACCACCGAGTGTGACTGCTGCTCACCG 4598
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
763 AGATCGTCCCGACTATCTATCTGTGTGTGTGCGCGGTGTTTCGCTGCTGCTGCGCG 822
QY 4599 GCTTCCTAGTTTCATGCTATTGGTCC 4626
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
823 TGTTCCTGGCGCACGCACTGATCGGCGC 850
```

```
RESULT 14
BE790435
LOCUS 601477188F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3880173 5',
DEFINITION mRNA sequence.
ACCESSION BE790435
VERSION BE790435.1 GI:10211633
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DFP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9646 row: o column: 22
High quality sequence stop: 564.
Location/Qualifiers
1. .669
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3880173"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_68"
/note="Organ: lung; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 165 a 172 c 186 g 145 t 1 others
ORIGIN
Query Match 1.7%; Score 102.2; DB 10; Length 669;
Best Local Similarity 50.1%; Pred. No. 6.4e-18;
Matches 311; Conservative 0; Mismatches 303; Indels 7; Gaps 2;
QY 1632 ACCTTGGGCTGCAACAGGATCCTCACCAGTACCTACCAAGAGCTCATTCGCGATG 1691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
25 ACCCTGGGCTCCCGACCTGGGAGTACCCCACTTGGCTGCTACAAGAGCTGATGATAC 84
QY 1692 TATGAAGCTGGGAGTGTCTATTCGAAGACTGCAAGGACTTCTTGTGGATGAATAGTG 1751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 TATAAGATGGGACCTGCTCTCTTAAATATGTGAAGACCTTCAACATGGATGAGTAGTG 144
QY 1752 GGACTAACCGTGCAGATGAACAGCTACTTTAAACCATTCGCAAGAGTTCACTGAC 1811
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 GCGCTTCTCTGAGACACCCCGGAGATTACCACCTCTCATGTGGACAACTTCTTCAAG 204
QY 1812 CACATCGACCTGTTGATGAAGAGTCTACAGCCGAGATGGTGCACACCTGATCATAC 1871
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 CACATTGACATCCACCCAGAAAACACCCACATCTTGGATGGGAATGCAGTCGACCTACAG 264
QY 1872 GAAGCAGCTGCAGATGATGAGCAAGATCGCTGCAGATCCGTTGAAGTTCAATCCTT 1931
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 GCAGAAATGTGTCGCTTTGAAGAGATCAAGGTCGAGTGGGATGAGCTATTGTT 324
QY 1932 GGCATCGCGG ---AAACGGCAGATCGCTTTTCAATTTGAACCATCATCTTCTGTGACGA 1987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
325 GGAGGCATCGGCCCTGTGACACATTCCTTCAACGAGCCAGGCTCCAGTCTGTTGCC 384
QY 1988 CTGACAAAGTCCAGCGCTGCACCCTTAAACTGTGGAGACACAGCTCGATTCTTCAA- 2046
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db      385 AGGACCGGTGTGAAGAGCGTGGCCATGGATACCATCTGGCCATCTAGGTTCTTCGAT 444
Qy      2047 --CACCATCGAAGAGTCCCAAGCCAGCCGTCACCCAGGGTTGGCACTTTGTCCCGC 2104
Db      445 GGAGAACTACCAAGTGGCCACCATGGCTTTGACGGTGGGGTGGGCACTGTCATGGAT 504
Qy      2105 GGCCTAAACATCGTGTGGTGGCACTGTGAAGGAAAGCCGACGCCATCCGCGAAGT 2164
Db      505 GCTAGAGAGGTGATGATCTTATCACAGGTGCTCACAGGCATTGCTGTACRAGGCC 564
Qy      2165 GTGGAAGGCCAGTACTGCTTTCGCCAGGTTCCATCTGTAGATGCACACATGCCA 2224
Db      565 ATCGAGGAGGAGTGAACACATGTGGACCGTGTCTGCTTCAGCAGCATCCCGCACC 624
Qy      2225 CCATCATCGTTGGATGAAGA 2245
Db      625 GTGCTGTGTGAAGAAGA 645

RESULT 15
BUI71522
LOCUS      847 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7928634 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6151420
            S, mRNA sequence.
ACCESSION BUI71522.1 GI:22685506
VERSION    BUI71522.1
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 847)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTP/DMP
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM3487 Row: k Column: 05
            High quality sequence stop: 536.
FEATURES   Location/Qualifiers
            1..847
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6151420"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_72"
                /notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                Average insert size 2 kb. Library constructed by Life
                Technologies."
BASE COUNT 219 a 210 c 226 g 190 t 2 others
ORIGIN
Query Match 1.7%; Score 102.2; DB 13; Length 847;
Best Local Similarity 49.9%; Pred. No. 7.4e-19;
Matches 311; Conservative 0; Mismatches 305; Indels 7; Gaps 2;
Qy      1632 ACCTGGGGTTCGACAGAGATCCTACCACTAGTACCTACCAAGAGCTCATTCGCATG 1691
Db      99 ACCCTGGGGTCCCGACCTGGGAGTACCCCACTTGGCTGTCTACAGRAGCTGATTGAATAC 158
Qy      1692 TATGAAGCTGGGGAAGTGTATCAAGACATGCAAGCATCTTGTGGATGAATACGTTG 1751
Db      159 TATAAGATGGGGACCTGTCTTTAATATGTGAGACCTTCACATGGATGATACGTG 218

```

```

Qy      1752 GGACTAACCCGTGAGGATGAAGACAGCTACTTTAAACACCTTCGCAAGAGTTCTCACTGAC 1811
Db      219 GGCCTTCTCGAGACCCACCCGGAGAGTTACCACCTCTTCATGTGACAACTTTTTCAG 278
Qy      1812 CACATCGACATCGTTGTATGAAGAGGTCTTACAGCCCGAGATGGTGCAAAACCCCTGATCCATAC 1871
Db      279 CACATTGACATCCACCCAGAAAACACCCACATTCTGGATGGGAATGCAGTCGACCTACAG 338
Qy      1872 GAAGCAGCTGCAGAGTATGAGGCAARGATCGCTGCAGAATCCGTTGAAGTTCAATCCTT 1931
Db      339 GCAGAATGTATGCGCTTTGAAGAGAAGATCAAGGCTGCAGGTGGGATCGAGCTATTGTT 398
Qy      1932 GGCATCGCGGG---AAACGGCACATCGCTTTTCATTGAACCATCATCTTCTCTGTACAGA 1987
Db      399 GGAGGCATCGGCCCTGTATGGACACATTGCCCTTCAAGGAGCCAGGCTCCAGTCTGTGTGTC 458
Qy      1988 CTGACAAAGGTCAGGCCCTGCACCCTAAACCTGTGGAGGACACGCTCGATTCTTCAA- 2046
Db      459 AGGACCCGTGTGAAGACGCTGGCCATGGATACCATCTCTGGCCAATGCTAGGTTCTTCGAT 518
Qy      2047 --CACCATCGAAGAGTCCCAACCCAGCCGTCACCCAGGGTTTGGGCACCTTTGTCCCGC 2104
Db      519 GGAGAACTCACCAAGGTGCCACCATGGCTTGACGGTGGGGTGGGCACTGTCATGGAT 578
Qy      2105 GCGCAAAACATCGTGTGGTGGCAACTGTGTGAAGGAAAGCGGACGCCATCCGCGGAAGT 2164
Db      579 GCTANAGAGGTGATGATCTCTTATCACAGGTGCTCAAGGCAATTTGCTCTGTACAGGCC 638
Qy      2165 GTGGAGAGGCCAGTCACTGCTTCTTCCAGGTTCCATCTGTAGATGCACAAACATGCCA 2224
Db      639 ATCGAGGAGGAGTGAACACACATGTGGACCGTGTCTGCTTCCACAGCATCCCCGCGACC 698
Qy      2225 CCATCATCGTTGGATGAAGCAGC 2247
Db      699 NGGTTTGTGTGACGAGGATGC 721

```

Search completed: October 1, 2003, 01:21:48  
Job time : 7510 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 19:02:07 ; Search time 259 Seconds  
(without alignments)  
10172.262 Million cell updates/sec

Title: US-10-019-284A-1  
Perfect score: 5969  
Sequence: 1 agtcgctgagccaccatt.....tgataatttgcgcgctgaa 5969

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata2/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata2/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata2/ina/FCIUS\_COMB.seq:\*  
6: /cgn2\_6/prodata2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3687	61.8	6911	1	US-08-311-174-4
2	172.2	2.9	3895	4	US-08-961-527-201
3	154.2	2.6	9769	4	US-08-961-527-30
4	139	2.3	732	4	US-09-134-001C-1277
5	124.8	2.1	714	4	US-09-107-532A-378
6	108.6	1.8	3615	1	US-08-920-812-17
7	108.6	1.8	3615	1	US-08-920-827-17
8	108.6	1.8	3615	1	US-08-921-177-17
9	108.6	1.8	3615	1	US-08-362-577C-17
10	108.6	1.8	3615	2	US-08-920-828-17
11	95	1.6	485	2	US-08-673-190A-3
12	92.4	1.5	1887	4	US-09-107-532A-1634
13	92	1.5	2295	4	US-09-221-017B-736
14	90.6	1.5	834	4	US-09-367-293-1
15	90	1.5	1830121	4	US-09-557-884-1
16	90	1.5	1830121	4	US-09-643-990A-1
17	84.4	1.4	4403765	3	US-09-103-840A-2
18	84.4	1.4	4411529	3	US-09-103-840A-1
19	73.8	1.2	1284	4	US-09-107-532A-1876
20	69.4	1.2	1752	4	US-09-107-532A-3157
21	65.6	1.1	357	2	US-08-673-190A-6
22	65.6	1.1	2726	4	US-09-634-238-33
23	65.6	1.1	2726	4	US-09-634-238-46
24	65.2	1.1	1034	4	US-08-367-293-2
25	64.4	1.1	1314	4	US-08-936-165A-153
26	62.2	1.0	8494	4	US-08-961-527-163
27	61.8	1.0	1164	4	US-09-252-991A-1543

c 28 61.8 1.0 1209 4 US-09-252-991A-1624  
c 29 61.8 1.0 2127 4 US-09-252-991A-1503  
c 30 58.2 1.0 4403765 3 US-09-103-840A-2  
c 31 58.2 1.0 4411529 3 US-09-103-840A-1  
c 32 58 1.0 325 4 US-08-961-527-351  
c 33 55.4 0.9 1437 4 US-09-134-001C-2228  
c 34 55 0.9 2435 4 US-09-634-238-131  
c 35 52.8 0.9 513 4 US-09-134-001C-297  
c 36 52.8 0.9 1971 4 US-09-107-532A-1429  
c 37 52.6 0.9 1846 4 US-09-634-238-183  
c 38 52.2 0.9 1896 4 US-09-107-532A-248  
c 39 51.4 0.9 28882 4 US-08-961-527-140  
c 40 50.4 0.8 744 4 US-09-134-001C-2556  
c 41 50 0.8 612 4 US-09-252-991A-1623  
c 42 50 0.8 936 4 US-09-252-991A-1504  
c 43 50 0.8 1296 4 US-09-252-991A-1670  
c 44 50 0.8 2061 4 US-09-252-991A-1544  
c 45 47.2 0.8 1344 1 US-08-844-010-1

## ALIGNMENTS

RESULT 1  
US-08-311-174-4  
; Sequence 4, Application US/083111174  
; Patent No. 5556776  
; GENERAL INFORMATION:  
; APPLICANT: TSUCHIYA, MAKOTO  
; APPLICANT: MIWA, KIYOSHI  
; TITLE OF INVENTION: SUCRASE GENE DERIVED FROM CORYNEFORM  
; TITLE OF INVENTION: BACTERIA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/311,174  
; FILING DATE: 23-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 046836/1992  
; FILING DATE: 04-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5556776man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 10-699-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6911 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-311-174-4

Query Match 61.8%; Score 3687; DB 1; Length 6911;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



|||||  
2161 AACTGTGGAAGGCCAGTCTGCTCTTCCAGGTTCACCTCTAGATGCACACAT 2220  
QY  
2221 GCCACCATCATGTTGATGAAGCAGCAGTATCCAAAGCTGGAAGACGCTGATCAGTACCG 2280  
Db  
2221 GCCACCATCATGTTGATGAAGCAGCAGTATCCAAAGCTGGAAGACGCTGATCAGTACCG 2280  
QY  
2281 TCTCATGAGCAATTAAGCTGGCTAGAACAAAAAGGAAAGTACTGTGGGGGTATG 2340  
Db  
2281 TCTCATGAGCAATTAAGCTGGCTAGAACAAAAAGGAAAGTACTGTGGGGGTATG 2340  
QY  
2341 CACACAGAACTTCCAGTTGCGCCCTGCGTACATGTACTCTCCGCGAGGCGAGCTC 2400  
Db  
2341 CACACAGAACTTCCAGTTGCGCCCTGCGTACATGTACTCTCCGCGAGGCGAGCTC 2400  
QY  
2401 AATGATCCCAACGGAATGTACGTGATGAGATACCCCTCCAGCTCTACTACCAGCAGAT 2460  
Db  
2461 CCAGGTTCCCTTCGACCAAGCGCAGCGGTGGGCTCACACACGACGCGGTTGACC 2520  
QY  
2521 CCAGGTTCCCTTCGACCAAGCGCAGCGGTGGGCTCACACACGACGCGGTTGACC 2520  
Db  
2521 CCAGGTTCCCTTCGACCAAGCGCAGCGGTGGGCTCACACACGACGCGGTTGACC 2520  
QY  
2581 GACCGGAGGATGCTATTCGGGTGGAGCCGCTATTTACTGACGCGCACACTTAACCTTTC 2640  
Db  
2581 GACCGGAGGATGCTATTCGGGTGGAGCCGCTATTTACTGACGCGCACACTTAACCTTTC 2640  
QY  
2641 TACACCGCAACCTTAAATAATTGACGGAAGCGCGCCGACCCAAACCTTGTGGAAGTC 2700  
Db  
2641 TACACCGCAACCTTAAATAATTGACGGAAGCGCGCCGACCCAAACCTTGTGGAAGTC 2700  
QY  
2701 GAGGACCAACTGGGCTGATGGGCGCATTCATCGCGTTCGCGCTAAATAATCCGCTTATC 2760  
Db  
2701 GAGGACCAACTGGGCTGATGGGCGCATTCATCGCGTTCGCGCTAAATAATCCGCTTATC 2760  
QY  
2761 GACGACCGCGCAGCGGTTTACACACCCGATTCACCGGATCCCATGATGATGATGATGATG 2820  
Db  
2761 GACGACCGCGCAGCGGTTTACACACCCGATTCACCGGATCCCATGATGATGATGATGATG 2820  
QY  
2821 GATGTTGGACATGTTCTTGGGGCCGACGCGGAAACCTCACCGGTGACGCGGTCTA 2880  
Db  
2821 GATGTTGGACATGTTCTTGGGGCCGACGCGGAAACCTCACCGGTGACGCGGTCTA 2880  
QY  
2881 TACCGCTCGACAGATCTTGAAGCTTGGGAATCTCCGCTGAAATCACCTTTGACCTCAGT 2940  
Db  
2881 TACCGCTCGACAGATCTTGAAGCTTGGGAATCTCCGCTGAAATCACCTTTGACCTCAGT 2940  
QY  
2941 GATGACAACTGTTCTGCTCTGATCTCTCCGATGCTGATGCTGATGCTGATGCTGATGCTG 3000  
Db  
2941 GATGACAACTGTTCTGCTCTGATCTCTCCGATGCTGATGCTGATGCTGATGCTGATGCTG 3000  
QY  
3001 AACCTTTTACGTTTCGGGATGAGAACTGGGAAATCTCGAGCTGCTGATTTCTGT 3060  
Db  
3001 AACCTTTTACGTTTCGGGATGAGAACTGGGAAATCTCGAGCTGCTGATTTCTGT 3060  
QY  
3061 CCACAAAGATTGGACCGCAATCCAGATGAGTTTACTACTACGCAAGCTCTGACCAAGTC 3120  
Db  
3061 CCACAAAGATTGGACCGCAATCCAGATGAGTTTACTACTACGCAAGCTCTGACCAAGTC 3120  
QY  
3121 GGATATGCTGCGCAACCTTGAAGGACGACCTTCCGCTTCCGAGGATTCAGCGAG 3180  
Db  
3121 GGATATGCTGCGCAACCTTGAAGGACGACCTTCCGCTTCCGAGGATTCAGCGAG 3180  
QY  
3181 CTGGATTTCGGCCATGAATTTACGACCGCAGGTTGCAAGTAAACGGTTCTGATGCTG 3240  
Db  
3181 CTGGATTTCGGCCATGAATTTACGACCGCAGGTTGCAAGTAAACGGTTCTGATGCTG 3240  
QY  
3241 CTCGTGGCTGATGGGCTGCGCCGCGCAGATGATCACCCCAACAGTTGCACAGGAAGA 3300  
|||||

Db  
3241 CTCGTGGCTGATGGGCTGCGCCGCGCAGGATGATCACCCCAACAGTTGCACAGGAAGA 3300  
QY  
3301 TGGGTGCACTGCTGACTGTGCGCCGCAAGCTTCAATTTGCGCAACACGCGATCTACCAA 3360  
Db  
3301 TGGGTGCACTGCTGACTGTGCGCCGCAAGCTTCAATTTGCGCAACACGCGATCTACCAA 3360  
QY  
3361 GAGCTCTTCTCCAGAGGGGAGTCGGGGTAAATCAGATCTGATTAAGTTCTCAACT 3420  
Db  
3361 GAGCTCTTCTCCAGAGGGGAGTCGGGGTAAATCAGATCTGATTAAGTTCTCAACT 3420  
QY  
3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTTCTGTG 3480  
Db  
3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTCCGTTTCTGTG 3480  
QY  
3481 GATCGTGATGATGATGCTGCTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 3540  
Db  
3481 GATCGTGATGATGATGCTGCTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 3540  
QY  
3541 GATAATACAGCATTGAGATAAATCTGAGCTGATGAGTGGGATTTTCATTCCTTTCCGGGC 3600  
Db  
3541 GATAATACAGCATTGAGATAAATCTGAGCTGATGAGTGGGATTTTCATTCCTTTCCGGGC 3600  
QY  
3601 CTTCAAGGTGACACTATTGAGAGATAAGTATATATAAAGGTCCTTTTGTGGGATTTGT 3660  
Db  
3601 CTTCAAGGTGACACTATTGAGAGATAAGTATATATAAAGGTCCTTTTGTGGGATTTGT 3660  
QY  
3661 ACAATACTTCGCAAAATCCCTTGATC 3687  
Db  
3661 ACAATACTTCGCAAAATCCCTTGATC 3687

## RESULT 2

US-08-961-527-201/c  
: Sequence 201, Application US/08961527  
: Patent No. 6420135  
: GENERAL INFORMATION:  
: APPLICANT: Charles Kunsch  
: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
: NUMBER OF SEQUENCES: 391  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Human Genome Sciences, Inc.  
: STREET: 9410 Key West Avenue  
: CITY: Rockville  
: STATE: Maryland  
: COUNTRY: USA  
: ZIP: 20850  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
: COMPUTER: HP Vectra 486/33  
: OPERATING SYSTEM: MSDOS version 6.2  
: SOFTWARE: ASCII Text  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/961,527  
: FILING DATE:  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Brookes, A. Anders  
: REGISTRATION NUMBER: 36,373  
: REFERENCE/DOCKET NUMBER: PB340P1  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (301) 309-8504  
: TELEFAX: (301) 309-8512  
: INFORMATION FOR SEQ ID NO: 201:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3895 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
US-08-961-527-201



QY	4379	CTGGTTAACGGCTACGACGTGGCGCCACCATGACCCGGCGGAGAAATGCCAAATGTGGTCC	4433
Db	9146	CTTCCAAACCTTTGGCAGTTGC-----TCAAGGTGGTGAAGTAAACGCGATGAAC	9096
QY	4439	CTGTTTGGTTTGGATGTTGCTCAAGCTGGTTACAGGGCACCGTGTCTCTGTGCTGGTG	4498
Db	9095	TTCCTTTGGTTT---CATCCCTGTTGTTGGTTTGAAGTTCCGTTCTTCTCCAGCCTTCATC	9039
QY	4499	GTCTCTTGGATTCTGGCAACGATCGAGAAGTTCTCTGCACAACGCACTCATGGGCACGTGCA	4558
Db	9038	ATCGGGGTCTCGGAGCTAAATTGAAAAAGCTGTCGCAAGTTGTTCCAGATGTC-ATT	8980
QY	4559	GACTTCTGATCACCCAGTGTGACTCTGCTGCTCACCGGCTTCCCTTAGTCTCATTTGCT	4618
Db	8979	GACCTCTTGCTAACACCATTCGTGACACTTTTGGTCATGCTATCTCCTTGGACTCTTTGTC	8920
QY	4619	ATTGGTCCACAAATGCGCTGGGTGGTGACTTGTGCGCACACGCTGTGCAGGAGCTCAT	4678
Db	8919	ATTTGGACCACTTTTCCACGTGTGTGAAGAACTACATCCTTATTGCTCAAAAGCGATTCTT	8860
QY	4679	GATTTCCGTGGTCCAGTCGGCGGTCTGCTTTTCGGTCTGGTCTACTCACCAATCGTTATC	4738
Db	8859	AGCATGCCATTGGTCTTGTTGGTTTCTGATTGGTGGGGTTTCCAAATTCATCGTCGTG	8800
QY	4739	ACTGGTCTGCACCACTCTCCCGCCCAATTTGAGCTGGAGCGTGTTCACACGAGG---TGGGA	4795
Db	8799	TCAGGTGTGCACCACACTTCTCAACTGCTTTGAAGTGCAAATTACTTGCTGTGACCATGCT	8740

[illegible]

Db 8100 TTGAAACAGAAATG-----CTGCTGAAGTTTGTATCATCTTGGTATGAT 8053  
QY 5516 ACCGTAACCTCAACGCGACGACCTTTAACCCGCTGAAGAAGCAGCGGATGAAGTCAA 5575  
Db 8052 ACTGTATCATGAACGCTGACGGTTTGAACAAAAAGTTGCTCAAGGTAATAAGGTGAA 7993  
QY 5576 GCAGGGAGCTGCTGTGTAATTCGATATGATGATGATGATGATGATGATGATGATGAT 5635  
Db 7992 GCTGGCGATGTTCTTGAACATTTGACTCAAAACAAATCGCTGACGCTGATGATGAT 7933  
QY 5636 ACCACGCCGATTTCTTTCGAATTAACAAGAAACCGGACCTGTAAACACACTTACCGTTG 5695  
Db 7932 ACAACAATGTTATCGTTTCAAAATACAGCTGACTAGCTTACAGTCCAGTCGCAACA 7873  
QY 5696 GCGCAATTTGAAGCGGAGGACCACTGCTCAACGTCGCGAAAGAA 5740  
Db 7872 GGTTCAGTTGTAAGGGGATGCTGTGATGATGATGATGATGATGATGATGATGAT 7828

RESULT 4  
US-09-134-001C-1277  
; Sequence 1277, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134, 001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1277  
; LENGTH: 732  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (20),(22),(32)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.

Query Match 2.3%; Score 139; DB 4; Length 732;  
Best Local Similarity 56.6%; Pred. No. 1.5e-33;  
Matches 306; Conservative 0; Mismatches 220; Indels 15; Gaps 2;

QY 4569 TCACCCAGTGTGACTCTGCTGCACCGGCTTCTTACGTTCAATGCTATGCTGTCGAC 4628  
Db 80 TGNCGCATATATCAATTTTATACAGCAATTAACATTTTATTTAGGCGCTG 139  
QY 4629 CAATGCGTGGGTGCTGCTGGCAGACGGTCTGACGAGCTCTATGATTTGCGTG 4688  
Db 140 TCACTGCTCAATAGTATTGTTGCTATCTGATGATGATGATGATGATGATGATGATG 199  
QY 4689 GTCCAGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4748  
Db 200 GAGCTATGCTGACTTATTTGTTTATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 259  
QY 4749 ACAGTCCTTCCCGCAATGAGCTG-----CAGCTGTTCACACAGGTTGAT 4796  
Db 260 ACCATATCTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 319  
QY 4797 CTTTCATCTTCGAACCGCATCCAGGCAATTCGCGAGGTCGAGGATTTTGGCAG 4856  
Db 320 CATTTATCTTCCCAATCGCAAGATGCTCAAAATTTGCAAGGTCGCTGAGGTTAGTG 379  
QY 4857 TGTTCCT---TCCTAGGCAAGAGTGAAGAGCTCAAGGCGCTTGCAGTGCTTCAGTGCT 4913  
Db 380 CATTCCTTATCATAGCAAAATATAAAATTAAGAGTGTTCCTCCGCGCGGTATTT 439

QY 4914 CCGCTGTTCTTGGTATTACAGAGCCTCGGATCTTCGGTGTGAACCTTCGGCTGCGCTGGC 4973  
Db 440 CAGCTTTACTAGGAATACAGAACCAATGTTTGGTGTCAATCTTTAAATTTGAGATATC 499  
QY 4974 GCTTCTACATTTGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTCGACTCTTTGATA 5033  
Db 500 CATTTATAGTGTCTGTTGCAGGATCAGGTATAGGTGGCGCTTATATTTCTTCAAG 559  
QY 5034 TCAAGGAGTGTGCTTGGCGCTGCAGGTTCTTGGGTGTGTTTCTTATGATGCTCCAG 5093  
Db 560 TAAAGCGATACGCTTGGTACAGCTGATACCTGATTTATATCTATAATACCTACAC 619  
QY 5094 A 5094  
Db 620 A 620

RESULT 5  
US-09-107-532A-378  
; Sequence 378, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 378:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 714 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1..714  
; SEQUENCE DESCRIPTION: SEQ ID NO: 378:  
US-09-107-532A-378

Query Match 2.1%; Score 124.8; DB 4; Length 714;  
Best Local Similarity 52.5%; Pred. No. 5e-29;



```
Db 2254 CCGCTGTTAATGCCCGCCATTATGCGCAGTTCGGGGCGGCTTCCTCTG 2195
Qy 4867 AGGGAAGATGAAGAGCTCAAGGCGCTTCAGGTGCTTCAGGTGCTTCGCTGCTTCTG 4926
Db 2194 CGAACCGGATGCGCAGAAAAAGTGTGCGGGGATCAGCGGCTTGACAGTCTGTTGG 2135
Qy 4927 TATTACAGAGCGCGCATCTCGTGTGAACCTTCGCTGCGTGGCGCTTCTACATGG 4986
Db 2134 TATCAGGACACGCGGTATATGGCTCAACCTGCGCGCTAGTACCCCTTTGTTATCG 2075
Qy 4987 TATCGTACGCGAGCTATCGGTGGCGCTTTGATGCACTCTTTGATATCAAGCAGTTCG 5046
Db 2074 CTGTATCAGTGGGCTTTGGGGCCACCATTTATGGCTACGGGCAACGAAAGTCTATC 2015
Qy 5047 GTTGGCGCTGCGAGGTTCTTGGGTGTTGTTTATGATGCTCCAGATATGTCATGT 5106
Db 2014 CTTGGTTGCGAAGTATTTACCTTCATGCAACCATCCCTCAACGGGAATGATTT 1955
Qy 5107 CTTGGTTGCGGCTAGTACCTTTGTCATCGCATTCGCGGAGCGATTTGCTTATGGCT 5166
Db 1954 CACCGTCTGGCGCAGGCTTA--TTGGCGGTGTCAATTGCGCATTTGCGCATTTGTCG 1898
Qy 5167 TTACTTGGTTCGCGCAAGCGCAGCATTTGATCCAGATGCAACCGCTCCAGTGCCTG 5226
Db 1897 TACGGTATGCTTCATTTTCATCCCGTAAACGCTAGCCAGC-----1856
Qy 5227 AGAAGCAACCAAGCGGAGCAGAGCAGCAGCGGAGAAATTTCAACGATTCACCATAT 5286
Db 1855 -GCAGGTGCGCGGAGAGAAACACAGAGGTTATTACACCACTGAGCAGGCGGTAT 1797
Qy 5287 CCAGGACCTTTACCGGCTGAAGCTATCGCAGTGAGCAGGCTCAGCGATGCCATGTTTC 5346
Db 1796 CTGTTCCAGATCAGCGGAGAGATTTGTCGCTCATTCAGCTGCTGATACACGTTTC 1737
Qy 5347 CAGCGGAAGCTTGGCTCAGGTGTTGCGATCGTCCCAACCAAGGCGCAGCTGTTTCCAC 5406
Db 1736 CAGTGGCTGTTGGTAAAGGTATTGCCATTCTGCGCTCGGTTGGTGAAGTGGTCTCG 1677
Qy 5407 AGTGAAGGGAAGATCGTGTGGCTTCCTCATCTGCTCAGCTTCGCGCTCGCACTAA 5466
Db 1676 GGTGCGGCTGAATTCCTGCTGTTGCGGCATTTACGCGCATTTG-----GCATGA 1623
Qy 5467 GGTGAGATGTTTCCAAATGATATGATATGATGCAATGTTTCGACACCGTAAACCT 5526
Db 1622 GTCAGATGATG-----TGTGGAGATCTGATTCATGTCGTTATCGACACCGTAAACT 1569
Qy 5527 CAACGCGACGACTTTAACCGCTGAGAGACGAGCGGATGAGTCAAAACGAGGAGCT 5586
Db 1568 GGCGGCAAAATCTTTTCCGCTCAGCTCAACGTGGGTGACAGGTCAATACAGCGATCG 1509
Qy 5587 GCTGTGTAATTCGATATGATGCAATTAAGGCTGCAAGTTATGAGTAAACACGCGAT 5646
Db 1508 GCTGATTTCTTTTCATATCCTGCTATTCGAGGCGGATTTGATCTGACGACGCGGT 1449
Qy 5647 TGTGTTTCGATTTACAGAAACCGGACCTGTAACACATTCAGGTTTGGCGCAATGA 5706
Db 1448 ATTAATCAGTAATAGCATGATTTTACGACGATTTATACCCCGGACGCGGAGATAAG 1389
Qy 5707 AGCGGAGCAACCTGCTCAACGTC 5731
Db 1388 CGCAGGTGAACCGCTGTTATCCATC 1364
```

## RESULT 7

US-08-920-827-17/c  
; Sequence 17, Application US/08920827  
; Patent No. 5770375

## GENERAL INFORMATION:

; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotugu  
; APPLICANT: Eda, Soji

TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920.827  
FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
STRAIN: Clinical Isolate EC-24  
US-08-920-827-17

Query Watch 1.8%; Score 108.6; DB 1; Length 3615;

Best Local Similarity 45.5%; Pred. No. 2.9e-23;  
Matches 666; Conservative 0; Mismatches 754; Indels 45; Gaps 6;

```
Qy 4270 TCGCGCGTTCGGGCTTCTTCCAGTGTGTTGGTTTACCGCAACCAAGCGTTTCGGTGG 4329
Db 2786 TCGCTTGTGTTTCTTCCCGATAATCCTGGATACACCGGGGAAACGCTTCGGCGA 2727
Qy 4330 CAATGATTCCTGGGCGCGGATGTTGATGCGATGTTCCCAACCCCTGGTTAAACGG 4389
Db 2726 TAATCCATTTACTGCATGGTGATTTGGTGGACGTTAGTGCATCCATTAATCTGTACTG 2667
Qy 4390 CTACGACGTGGCGCCACCATGACCGCGGGCGGAAATGCCAATGTGGTCCCTGTTGGTTT 4449
Db 2666 CTTTTCGAGACGCGCAAAA-----CGGATCGCTGGGGCTGGATTTCTCGGTAT 2615
Qy 4450 GGATGTTGCTCAAGCTGTTTACAGGCGACCGTCTCTCTGCTGCTGGTGGTCTCTGGAT 4509
Db 2614 TCGGTGCATGTTGGAATTTACTGTCATCGCTGATTCCTCATATTTTCTGCTGCTGTT 2555
Qy 4510 TCTGGCAACGATCGAGAAGTTCTCTGCAAGCGACTCATGGCGACTGCAGACTTCTCTGAT 4569
Db 2554 GTGCAGCATTCGGAACGCGGACTTAATGCGTGGTTACGCTCGGCAATCAAAAATTTCTT 2495
Qy 4570 CACCCAGTGTGACTCTCTGCTCAGCGCTTCCTTAGCTTCATGCTATTTGGTCCAGC 4629
Db 2494 CACACCATTTGCTATGCTGATGTTATACACCCGCTACCTTTCTGCTGGTGGGCGCGT 2435
Qy 4630 AATGCGCTGGTGGGTGACTTCTGCGCACACGCTCTGCGAGGACTCTATGATTTTCGGTGG 4689
Db 2434 ATCAACCTGATAAGCGAACTGATTCGCCCGCGTTATCTCTGCTTTATCAGCGCGTCTCC 2375
```

Qy	4690	TCAGTGC	GGGGTCTGCTTTTCGGTCTGTGCTACTACCAAAATCGTTATACACATGGTCTGCA	4749
Db	2374	TGCAATTTGGGGCGCGGTAAATGGGCGCTTTTGGCAAATCTTCGTCAATGTCGAGATGCA	2315	
Qy	4750	CCAGTCCCTTCCGCCCAATTCAGCTGGAGCTGTTCAAACCAGGGTGGATCCTTCATCTTCG	4809	
Db	2314	CTGGGCGCTGGTGC	CGGTGTATCAATAACTTCACCGTGTGGGCTACGACACCATGAT	2255
Qy	4810	AACCGCATCCAGTGG---GCCAAATATCCGGCAGGGTGCAGCATGTTTGGCGAGTGTCTTCTCT	4866	
Db	2254	CCGCGTGTAAATGCCCGCCATATATGCGCAGGTCGCGGGCGGCTCGGGCTCTTCTCTCTG	2195	
Qy	4867	AGCGAAGATGAAAAGCTCAAGGCCCTTCAGAGTCTTCAGGTGTCTCCGCTGTCTTCTGG	4926	
Db	2194	CGAACCGATGCGCACA	AAAAAGTGTGTGGGGATCAGCGCGTTCAGCAGTCTGTCTGG	2135
Qy	4927	TATTACAGACCTGCGCATCTTCGGTGTGAACCTTCGCCCTGCCGTGGCCGTTCACATTCG	4986	
Db	2134	TATCACCGAACACGCGGTATATGGCTGCAACCTGCGCGTAAGTACCCCTTTGTATATCG	2075	
Qy	4987	TATCGGTACCGCAGCTATCGGTGGCCCTTTGATTGCACCTCTTTGATATCAAGGCACTTGC	5046	
Db	2074	CTGTATCAGTGGGCTTTGGGGCCACCATTAATTTGGCTACGGCAAAAGTCTACTC	2015	
Qy	5047	GTTGGCGCTGCAGTCTTCTTGGGTGTGTTTCTATTTGATGCTCCAGATATGTCATGTT	5106	
Db	2014	CTTTGGTTTTCGCAAGTATTTTACCTTCATGCAAAACCATCCCGTCAACGGAAATGATTT	1955	
Qy	5107	CTTTGGTTTGGCGGTAGTTACCTTTGTCATCGCATTCGGCGCAGCATTCGTTATGGCCT	5166	
Db	1954	CACCGTCTGGGCGCAGGTTA---TTGGCGGTGTCAITTGCCATCGGTTGCGCATTTGCGG	1898	
Qy	5167	TTACTTGGTTTCGGCGCAACGGCAGCATTTGATCCAGATGCAACCGCTGCCGTCCAGTGCCTGC	5226	
Db	1897	TACGGTGTATCTTCATTTCAATCACCGCTAAACGTACGCCAGC-----	1856	
Qy	5227	AGGAACGACCAAGCCGCAAGCAGAACGACCGCAGAGATTTTCAAGAGGATTCACCAATCAT	5286	
Db	1855	-GCAGGTGCCCCGAGAGAGAAAACACAGAGGTTATTACACCACCTGAGCAGGGCGGTAT	1797	
Qy	5287	CCAGGCACCTTTACCCGGTCAAGCTATCGCACTGACGAGGTGACGCGATGCCATGTTTC	5346	
Db	1796	CTGTTCACCGATGACGGGAGAGATTTGTTCGCTCATTCAGTCTGCTGATACCCAGTTCG	1737	
Qy	5347	CACGCGAAAGCTTTGGCTCAGGTGTGGATTCGCTCCGCCACCAAGGCGCAGCTGTGTTTCA	5406	
Db	1736	CAGTGGCCTGTTGGGTAAAGGTATTGCCATTCCTGCCCTCGGTTGGTGAAGTGGTCTCC	1677	
Qy	5407	AGTCAGCGGAAGATCGTGTGGCTTCCCATCTGGTTCAGCGCTTCGCACTCCGCACTAA	5466	
Db	1676	GGTTGGGGTCTGAATGCTCTGTTTTCGCCACATTAACAGCCATTTG-----GCATTTGA	1623	
Qy	5467	GGCTGAGGATGTTTCCAAATGTGATATCTTGTATGCACATTTGGTTTCGACACCGTAAACCT	5526	
Db	1622	GTCAGATGATGG-----TGTGGAGATCTCTGATTCATGTCCGTATCGACACCGTAAACCT	1569	
Qy	5527	CAACGGCACGCACCTTAAACCGCTGGAAGACAGGGGGATGAAGTCAAAACGACGGGAGCT	5586	
Db	1568	GGACGGCAAAATCTTTTCCGCTACGCTCAACGTGGGTGACAAAGGTCAATACAGCGCATCG	1509	
Qy	5587	GCTGTGTAATTCGATATTCGCAATTAAGGCTGCAAGTTATGAGGTATACCAACGCGCAT	5646	
Db	1508	GCTGATTTCTTTGATATTCCTGCTATTCGCGAGCGCGGATTTGATCTGACGACGCGCGT	1449	
Qy	5647	TGTTGTTTTCGAATTAAGAAGAAACCGGACCTGTAAACACATTCAGGTTTGGCGAAATTTGA	5706	
Db	1448	ATTAAATCAGTAAATAGCAGATGATTTACGGACGATTAACCCACGCGCAGCGCAGATAG	1389	
Qy	5707	AGCGGGACCCAACTGCTCAACGTC	5731	
Db	1388	CGCAGGTGAACCGCTGTTATCCATC	1364	

RESULT 8  
 US-08-921-177-17/c  
 Sequence 17, Application US/08921177  
 Patent No. 5798211  
 GENERAL INFORMATION:  
 APPLICANT: Ohno, Tsuneya  
 APPLICANT: Matsuhisa, Akio  
 APPLICANT: Uehara, Hirotsugu  
 APPLICANT: Ego, Soji  
 TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION NUMBER: US/08/921,177  
 FILING DATE: 29-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/362,577  
 FILING DATE: 27-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rin-Laures, Li-Hsien  
 REGISTRATION NUMBER: 33,547  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3615 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: Clinical Isolate EC-24  
 US-08-921-177-17

## RESULT 8

QY	4570	CACCCAGGTGTTGACTCTGCTGCTCAGCCGGCTTCCTTACGTTCAATGCTATTGGTCCAGC	4629
Db	2494	CACACCATGCTATGCTGATGGTTATCACACCGTCACTTCTGCTGGTGGGGCGCT	2435
QY	4630	AATGGCTGGTGGGTGACTTCTGCTGCGCACACGGTCTGCAGGGACACTATGATTTCGGTGG	4689
Db	2434	ATCAACCTGGATAAGCGAACTGATTGCGCGGGTATCTCTGGCTTTATCAGCGGTTCC	2375
QY	4690	TCAGTCGGGGTCTGCTTTTCGGTCTGCTACTCACCAATGGTTATCACTGGTCTGCA	4749
Db	2374	TGCATTGGGGCGCGTAATGGCGGCTTCTGGCAAACTCTCGTCATGTTGGAGCTGCA	2315
QY	4750	CCAGTCTCTCCGCCAAATGAGCTGGAGCTGTTCACCAAGGTGGATCCTTCATCTTCG	4809
Db	2314	CTGGGGCTTGGTCCGGTGTGTATCAATTAACITCACCGTGTCTGGGTACGACACCATGAT	2255
QY	4810	AACCGCATCCATG - --GCCAATATCGCGAGGGTCAGCATGTTTGGCAGTGTCTTCCT	4866
Db	2254	CCCGCTGTTAAATGCCCGCCATTATGCGCAGGTTCGGGGGGCGCTCGGGCTCTCTCTG	2195
QY	4867	AGCGAAGATGAAAGCTCAAGGCGCTTCAGGTGCTTCAGGTGTCCTCGCTCTTCTTGG	4926
Db	2194	CGAACCGATGCGCAAAAAAGTGTGGCGGATCAGCGCGTTGACGAGTCTGTTTGG	2135
QY	4927	TATTCAGAGCCTGCGATCTTCGGTGTGAACCTTCGCCCTGCGCTGGCGGTCTACATTTG	4986
Db	2134	TATCAACGAACACAGCGGTATATGGCTCAACCTGCGCGCTAAGTATCCCTTTGTTATCG	2075
QY	4987	TATCGGTACGCGAGCTATCGGTGGCGCTTTGATTGCACCTCTTTGATATCAAGGCATTC	5046
Db	2074	CTGTATCATGTTGGGCTTTGGGGGCCACCATATTGGCTAGCGCAACGAAAGTCTACTC	2015
QY	5047	GTTGGCGCTGCAAGTTCCTTGGGTGTGTTCTATTGATGCTCCAGATATGTCATGTT	5106
Db	2014	CTTTGGTTTGGCAAGTATTTTCACTTCATGCAAAACCATCCCGCTCAACGGAAATTCATTT	1955
QY	5107	CTTGGTTTGGCGGTAGTTACCTTTGTCATCGCATTCGGCGAGCGATTGCTTTATGCGCT	5166
Db	1954	CACCGTCTGGCGCAGGTTA - --TTGGCGGTGTCATGGCATCGTTTGGTCATTTGTCG	1998
QY	5167	TTACTTTGGTTTCGCCAAGCGCAGCATTTGATCAGATGCAACCGTCTCCAGTGCTGTC	5226
Db	1897	TACGGTGATGCTTCATTTCATCACCGCTAAACGTCAAGCCAGC-----	1856
QY	5227	AGCAAGCAACAAAGCCGACGACGAAGCACCCGAGAAATTTTCAACGATTCACCATCAT	5286
Db	1855	-GCAGGTGCCCCGAGAGAGAAACACCAAGGTTATTACACACCTTGAGCAGGGCGGTAT	1797
QY	5287	CCAGGACACCTTTGACCGGTGAAGCTATTCGCACATGAGCAGGTGTCAGGGATGCCATGTTTC	5346
Db	1796	CTGTTCAACCGATCAGCGGAGAGATTGTGCGCTCATTCACGTGCTGATACCACTGTTGC	1737
QY	5347	CACGGNAAGCTTGGCTCAGGTGTTGGATCTGCCACCAAGGGGCGAGTGTGTTTCAAC	5406
Db	1736	CAGTGGCCTGTTGGGTAAAGGTATTGGCATTCGCGCTCGGTTGGTGAAGTGGTCTTCTCC	1677
QY	5407	AGTGAGCGGAAAGATCGTGTGGCGCTTCCCATCTGTGTCAGCGCTTCGACCTCGCGCAATA	5466
Db	1676	GGTTGGGGTTCGAATGCTTCGTGTTTGGCCACATPACAGCCCATG-----GCATTGA	1623
QY	5467	GGCTGAGGATGTTTCCAAATGTGATATCTTTGATGACATTTGGTTTCGACACCGTAAACCT	5526
Db	1622	GTCAAGATGATG-----TGTGGAGATTCCTGATTCATGTGCGTATCGACACCGTAAACT	1569
QY	5527	CACGGCACGCACCTTTAAACCGGTGAAGAAGCAGGGCGGTGAAGTCAAAACGACGGGAGCT	5586
Db	1568	GGACGGCAAAATCTTTTCGCTCAGCTCAAGTGGGTGACAAGGTCAATACAGCGATCG	1509
QY	5587	GCTGTGTGAATTCGATATTGATGCCATTAAAGCTGACGGTTATGAGGTATACCAAGCCGAT	5646
Db	1508	GCTGATTTCTTTGATATCCCTGATTTCTGCGAGCGCGGATTTGATCTGACGACGCGGT	1449

QY 5647 TGTGTTTCGAATTACAAGAAACCGGACCTGTAAACACATACCGTTTGGCGGAATGA 5708  
 Db 1448 ATTAATCAGTAATAGCGATGATTTTACGGACGTATTACCCACGCGCACGGCGCAGATAAG 1389  
 QY 5707 AGCGGGAGCCAACTGCTCAACGTC 5731  
 Db 1388 CGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 9  
 US-08-362-577C-17/c  
 ; Sequence 17 Application US/08362577C  
 ; Patent No. 5807873  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/362,577C  
 ; FILING DATE: 27-MAR-1995  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3615 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Escherichia coli  
 ; STRAIN: Clinical Isolate EC-24  
 ; US-08-362-577C-17

Query Match	1.88;	Score 108.6;	DB 1;	Length 3615;
Best Local Similarity	45.5%;	Pred. No. 2.9e-23;		
Matches 666;	Conservative 0;	Mismatches 754;	Indels 45;	Gaps 6;
Qy	4270	TGGGCCCTTCGCTCTTGGCCAGTGTGGTTGTTTCCACGCACCAACGAGGTTTCGTGG	4329	
Db	2786	TGCCTGTTTGGTTCCTCCGATAAATCTGGGATACACGCGGGGAAGCCTTCGCGA	2727	
Qy	4330	CAATGAGTCTCTGGCGCGCGSCATTGATGGCGATGGTTCCCAACCTGFTTAACGG	4389	
Db	2726	TAATCCATTACTGCCATGTGTGTTGGAGCGTTAGTGCATCCATTAATCTGTACTG	2667	
Qy	4390	CTACGACGTGGCGCGCCACCATGACCGCGGCGGAATCCCATGTGCTCCGTGTTGGTTT	4449	
Db	2666	CTTTTCGAAACGGCGAAAA-----GCGGATCGCGCTGGGGCTGGATTCCTGGGTAT	2615	
Qy	4450	GGATGTTGCTCTCAAGCTGTTTACCAGGCGACCGCTCTTCCTGTGCTGGTGTCTCTCGAT	4509	

Query Match	1.8%	Score 108.6	DB 1	Length 3615
Best Local Similarity	45.5%	Pred. No. 2.9e-23		
Matches 666	Conservative 0	Mismatches 754	Indels 45	Gaps 6
Qy	4270	TGGGCCCTTCGCTCTTGGCCAGTGTGGTTGGTTTCCACGCACCAACGAGGTTTCGTGGG	4329	
Db	2786	TGCCTGTGTTTGGTTCCTCCCGATAATCTCTGGGATACACGCGGGGAAGCCTTCGCGA	2727	
Qy	4330	CAATGAGTCTCTGGCGCGCGGCATTTGTTATGGCGATGGTTTCCCAACCTGTTTAACGG	4389	
Db	2726	TAATCCATTACTGCCATTGTGATTGGTGAGCGCTTAGTGCATCCATTAAATCTGTACTG	2667	
Qy	4390	CTACGACGTGGCGCGCCACCATGACCGCGGGCGGAATCCCATGTGCTCCCTGTTTGTTTT	4449	
Db	2666	CTTTTCGAAACGGCGCAAAA-----GCGGATCGCGCTGGGGCTGGATTCTCTGGGTAT	2615	
Qy	4450	GGATGTTCTGCTCAAGCTGTTTACCAGGCGACCGCTCTTCCCTGTGCTGGTGTCTCTTCGAT	4509	

Db 2614 TCGGGTACATTTGTTGAATTAATCTGCTCATCGGTTATTCCTCATTTATTTTCTCCCTGGTT 2555  
QY 4510 TCTGGCAACGATCGAGAGTTCTGCAACAGCAGCTCATGGGACATGCGACACTTCTGTAT 4569  
Db 2554 GTGACGATTTCTGGAACCGCATTTAATGCGTGTACCGTCCGCAATCAAAATTTCTT 2495  
QY 4570 CACCCAGTGTGACTCTGTCTGCTACCGGCTTCCCTTACGTTCAATGCTATTTGGTCCAGC 4529  
Db 2494 CACACCATTTGCTATGCTGTATGTTATACACCGCTCACCTTTCTGCTGTGGGCGGCT 2435  
QY 4630 AATCGCTGGTGGTGTGCTGCTGCGACAGCTCTGCGAGGACTCTATGATTTTCGGTGG 4689  
Db 2434 ATCAACCTGGATAGCGACTGATGCGCGGTATCTCTGCTTTATCGCGGTTC 2375  
QY 4690 TCCAGTCGGGCTGTGTTTTCGGTCTGCTCTACTACCAATCGTTATCACTGGTCTGCA 4749  
Db 2374 TGCATTTGCGGCGCGGTAAATGCGCGCTTTCGCAAAATCTCTGCTATGTTCCGACTGCA 2315  
QY 4750 CCAGTCTTCCGCGCAATGAGCTGGAGCTGTTCAACACGAGGTGGATCCCTCATCTTCGC 4809  
Db 2314 CTGGGCGTGTGCGGCTGTGTAATCAATCTTCAACGCTGTGGCTTACGACCATGAT 2255  
QY 4810 AACCGCATCCATG---GCCAATATCGCGAGGTGCGACATGTTTGGCAGTGTCTTCTCT 4866  
Db 2254 CCAGCTGTAATGCCGCGCAATATGCGCGAGGTGCGGCGCTCGCGCTCTTCTCTCTG 2195  
QY 4867 AGCAAGAGTGAAGCTCAAGGCGCTTGCAGGTGCTTCAAGTGTCTCCGCTGTCTTGG 4926  
Db 2194 CGAAGCGGATCGCGAGAAAGTGGTGGGCGATCAGCGCGTGTGACGAGTCTGTTGG 2135  
QY 4927 TATTACAGAGCTGCGATCTTCTGGTGTGAACCTTCGCTGCGCTGCGCGCTTCTACATGG 4986  
Db 2134 TATCAGCAACGCGGTATATGCGTCAACCTCGCGTAAGTACCCCTTGTGTTATCGC 2075  
QY 4987 TATCGGTACCGCAGTATCGTGGCGCTTTCATTGCGACTCTTTGATATCAAGGAGTGC 5046  
Db 2074 CTGTATCAGTGGGCTTGGGCGCCACCATTTATGGTACGCGCAACGAAAGTCTACTC 2015  
QY 5047 GTTGGGCTCGAGTCTTCTGGTGTGTTCTTCTATTTGATCTCCAGATATGTCATGTT 5106  
Db 2014 CTTGGTTGCCAGTATTTTCACTTCATCCAAACCATCCCGTCAACGGGAATGATTT 1955  
QY 5107 CTGTGTTGGCGGTAGTTACTTGTGTCATCGCATTCGGCGCAGGATGTTGTTATGGCT 5166  
Db 1954 CACCGTCTGGCCAGCGTTA---TTGGCGGTGTCTATTTGCCATCGGTTGCGCAATTCGG 1898  
QY 5167 TTACTTGTTCGCGCAACGCGCAGTATGATCCAGATCGAACCGCTGCTCCAGTGCCTGC 5226  
Db 1897 TACGGTGTGTTCTATTTTCACTACCGCTTAACGCTACGCCAGC----- 1856  
QY 5227 AGGAACGACCAAGCCGAGAGAGACACCGCGAGATTTTCAACAGATTTCCACCATCAT 5286  
Db 1855 -GCAGGTGCCCCGAGAGAAACACACAGAGGTTATTACACCACTGACGAGGCGGTAT 1797  
QY 5287 CCAGCAGCTTTGACCGTGAAGTATCGCACTGAGCAGCGTGCAGGATGCATGTTTGC 5346  
Db 1796 CTGTTACCGGATGAGGAGAGATGTTGTCGCTCATTCACGTCGCTATACCACTTGC 1737  
QY 5347 CAGCGGAAGCTTGGCTCAGTGTGCGATCTCCACCAAGGGGCGAGTGTGTTTACC 5406  
Db 1736 CAGTGGCTTGGGTAAAGTATTTGCCATCTTGCCTCGGTGTTGTAAGTGCCTTCTCC 1677  
QY 5407 AGTGAGCGGAAGATCGTGGTGGCTTCCCATCTGGTCAAGCTTTCGAGTCCGCATAA 5466  
Db 1676 GGTTCGGGTCGAATTTGTTTGTTCGTCACATATCACGCCATG-----GCATTGA 1623  
QY 5467 GCGTGAAGTGTTCCTCAATGTTGATATCTTGTATGACATTTGTTTCGACACCGTAAACCT 5526  
Db 1622 GTACAGATAGG-----TGTGGAGATCTCTGATTCATGTCGTATCGACACCGTAAACT 1569  
QY 5527 CAACGCGCAGCACTTTAAACCCGCTGAAGACGAGGCGGATGAGTCAAGCAGGGGAGCT 5586

Db 1568 GGACGGCAAAATCTTTTCGGCTCAGCTCAACGTGGTGACAGGTCATACAGGCGATCG 1509  
QY 5587 GCTGTGTGAATTCGATATTGATGCCATTAAAGCTCGAGGTATAGGTAACACCGCGAT 5646  
Db 1508 GCTGATTTCTTTGATATCCCTGCTATTCCGCGAGCGGATTTGATCTGACGACCGGT 1449  
QY 5647 TGTGTTGTCGATTTACAAGAACCGGACCTCTTAACACTTACGTTTGGGGGAAATTTGA 5706  
Db 1448 ATTAATCATGTAATAGGATGATTTTACGACGCTATATACCCACGCGGCGCAGTAAG 1389  
QY 5707 ACGGGAGCAACCTGCTCAAGTC 5731  
Db 1388 CGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 10  
US-08-920-828-17/c  
; Sequence 17, Application US/08920828  
; Patent No. 5853998  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,828  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-5300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3615 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-24  
; US-08-920-828-17

Query Match 1.88; Score 108.6; DB 2; Length 3615;  
Best Local Similarity 45.5%; Pred. No. 2.9e-23;  
Matches 666; Conservative 0; Mismatches 754; Indels 45; Gaps 6;  
QY 4270 TGCCTGCTCGCTTCTGTCAGTGTGTTTTCACGCAACCAAGCGTTTCGGTGG 4329  
Db 2786 TGCCTGTTTGTCTCTCCGATATCTCCGATATCTCCGCGGGAACGCTTCGCGCA 2727



Db 429 CGGGGGGGTTTCCACACCATGAATTTCTTCGGACTCGATATCGCOATGATCGGCTATCA 370  
Qy 4474 GGGACCGCTGCTTCTGTGCTGTGCTCTTCTTGGATTCTGGCAACGATCGAAGATTCT 4533  
Db 369 GGTACGGTGTCCGGTGTGCTGACGGTGTGTTATGACGGTCTGGAACAGCT 310  
Qy 4534 GCACAGCGACTCATGGGCACTCCAGACTTCCTGTATCACCCAGTGTGACTCTGCTGCT 4593  
Db 309 GCGCAGAGTATGCCAACGCGTGGACCTGATCCTGACGCCATTCTGACGGTGTGAT 250  
Qy 4594 CACCGGCTTCTTACGTTTCACTTATGTTGCTCAGCAATGCGTGGTGGTGGTACTTGT 4653  
Db 249 CTCCGGCTGCTGCCATGCTGTTTATCGCCCGCGCGGCGACACTGGTGGACGAT 190  
Qy 4654 GGCACAGCGTCTCAGAGACTCATGATTTCCGTGTGCTCAGTCCGGGGTCTGCTTTTCGG 4713  
Db 189 CTCCTTGTGCTCAGCAGCTGATGCCCCACGCGGTGTTGTCGCCGGATTCTGTTCCG 130  
Qy 4714 TCTGTTCTACTCACCATTCTTATCACTGTGCTGCACAGTCTCTCCCGCAATTGAGCT 4773  
Db 129 CGGCCGTATTCGCCATCGTCATCCCGCATCCACAGCTTCCACGCCGTAGAGG 70  
Qy 4774 GGAGCTGTTCAACCA 4788  
Db 69 GGGGCTGCTCGGCA 55

RESULT 12  
US-09-107-532A-1634  
; Sequence 1634, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1634:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1887 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO

ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1887  
SEQUENCE DESCRIPTION: SEQ ID NO: 1634:  
US-09-107-532A-1634

Query Match 1.5%; Score 92.4; DB 4; Length 1887;  
Best Local Similarity 44.7%; Pred. No. 2.6e-18;  
Matches 638; Conservative 0; Mismatches 736; Indels 54; Gaps 5;

Qy 4235 ACGGGTGTGTGAGATGATCAACCTGATGGCATCTGCGCGTTCGCGTTCCTGCGCAGTG 4294  
Db 430 ACGGTACTTATCAATATTGTTGCTGCAGCAGATGGTCTCTTACCTTCTTACCTGTC 489  
Qy 4295 TTGGTTGGTTTACCGCAACCAAGGTTTCCGTCGCAATGATGTTCTTGGGCGCGGATTT 4354  
Db 490 ATGCTTGTCTTTACAGCAGCTTAAAGTTCAAAGCAACCAAGTTTATGACGATGCAATT 549  
Qy 4355 GGTATGGGATGGTGTTCACACCTGTTTACGCTACGAGTGCGCGCCACCAATGACC 4414  
Db 550 GGTATGGCAGTGTGTATCCAGCTATCACAGAGTTGCTGGTGGGAGAGCTATCAGT 609  
Qy 4415 GCGGCGAAAGCCCAATGTGGTCCCTGTTTGGATGTTGCTCAAGCTGGTTACCAG 4474  
Db 610 TTCTTTGGGATCCCACTGAT-----TCCTTCCACCATCTGTTTACT 651  
Qy 4475 GGCACGGTCTTCTGTGCTGTGCTCTTGTGATTCTGGCAACGATCGAAGTTCCTG 4534  
Db 652 TCTTCTGTCATACCAATCATCTAGCGTATGGTTCAAAGTAAATTAAGCCATTCGTA 711  
Qy 4535 CACAAGCGACTCATGGCAGCTCAGACTTCTGTATCACCCAGTGTGACTCTGCTGTC 4594  
Db 712 AAAAAAGTATTCCGCAATTTCTGCAATGATTTAGTTCGCTGTGCTTTAGTAGTG 771  
Qy 4595 ACCGGTCTTCTTACGTTCTTCTATGTTGTCAGCAATGCGCTGGTGGTGGTGGTGGT 4654  
Db 772 ATGTTTCCATACATTTCTGCTTAGACCAATCGGAGCGGTTGCCGGTAATGCTTTA 831  
Qy 4655 GCACACGGTCTCAGGAGCTCATGATTTGCGTGTGTCAGTGGCGGGTCTGCTTTTCGT 4714  
Db 832 GCGGGTCTCTCAACTCGATTTATGATTTAGTTCCTCAATCGTAGCAGGCTTATCATG 891  
Qy 4715 CTGGTCTACTACCAATCGTTATCTACTGTCTGACCACTCTTCCCGCAATTCAGCTG 4774  
Db 892 AGTTTATGCAAGTATTCGTGATGTCGGCATGCTATGGGATTCGTACCGCATCATGTT 951  
Qy 4775 GAGCTGTTCAACCAAGGTTGGAT---CCTTCATCTTCGCAACCGCATCCATGCCAATATC 4831  
Db 952 TTGAATATCGAACAATATGGTTTGTATGATGATGATGATGATGATGATGATGATGAT 1011  
Qy 4832 GCGCAGGTTGACAGATGTTTGGCAGTGTCTTCTTAGCAGAGAGTGAAGCTCAAGGC 4891  
Db 1012 GCACAAAGTGGGCTCTTCTAGCTGATGATTCGGAACAAAAGATACGAACTTCGAGA 1071  
Qy 4892 CTGTCAGGTGCTTCAGGTGCTCCGCTGCTTGTGTTGTTATACAGAGCTTCGATCTTCGT 4951  
Db 1072 TTAGGATTTCTTCTACAGTACTGTTTATTTGATCCTGATCCTGATGATGATGAT 1131  
Qy 4952 GTGAACCTTCGCGTGGCGTGTTCATGTTGATCGGTATCGGTACCGCACTATCGGTGC 5011  
Db 1132 GTGACATTACCTTTGAAGAAACCATTCATTGCGGCTTGTATTCTTGGAGGATTTGGTGA 1191  
Qy 5012 GCTTTGATTCAGCTCTTGTATCAAGGAGTTCGTTGGCGCTCAGGTTCTTGGGT 5071  
Db 1192 GCTATTATCGGATTTCCGGAGTAAAGCTTTCTCCAGCAGTTTGGTCACTTATTCAGC 1251  
Qy 5072 GTTGTCTTCTATGATGCTCCAGATATGGTATGTTCTTGTGTTTGGCGGTAGTACCTTT 5131  
Db 1252 ATTCCGACITTCATTAAATACGGTGGATGTTGATAGAGTCAAAATGTCAGACAGTAA 1311  
Qy 5132 GTCATCGCATTCGGCGAGGATGCTTATGGCTTTACTTGGTTCCGCGCAACGCGCAGC 5191

Db 1312 GCTACAGGAATGCTTTGTTCTGCTAGTACATTTAGGATTTGACGAA 1371  
QY 5192 ATTGATCCAGATCAACCGCTGCTCCAGTCCCTGAGGACGACCAAGGAGGAA 5251  
Db 1372 CAACACAGCAAT-----CAATTAGAAACAACATGCGAATGACGA 1416  
QY 5252 GCACCCGAGAAATTTCAACAGATTCACCATCATCCAGGACCTTTGACCGGTGAAGCT 5311  
Db 1417 GAACCGATTACTTCAGACGGCAT-----ACCTTAAAGTCTTTAAAGGTAAGTA 1470  
QY 5312 ATGCACTGAGCAGCTGACCGATGCCATGTTTCCAGCGGAAGCTTGGCTCAGCTGT 5371  
Db 1471 CTGCCATTGAGTGAAGTACAGACCAAGTCTTTCTCTGCGGTGATGGAAAGTATC 1530  
QY 5372 GCGATGCTCCCAAGGAGGAGCTGTTTCCAGGATGAGCGGAAGATCGTGTGGCC 5431  
Db 1531 GCTATTGATCCAGAGTAGGAGAACTAGTTCACCTGCAGACGAGAGATCAGACTATT 1590  
QY 5432 TTCCATCTGCTCAGCTTTCGAGTCCGACCTAAGCTGAGGATGTTCCATGTGAT 5491  
Db 1591 TTCCAACTGACATGCTGTAGGATCACACGA-----CTGACGAGCGGAG 1638  
QY 5492 ATCTTGATGACATTTGTTTCGACACCGTAAACCTCAACGGCAGGACCTTTAACCGCTG 5551  
Db 1639 ATTTGATTCATATCGGGATGGATACAGTAGAATTAATGGGAATGGCTTTGAGATCTTA 1698  
QY 5552 AAGAAGCAGGCGATGAAGTCAACGAGGAGCTGCTGTGTGAATTCATGATTTGATGCC 5611  
Db 1699 GTCAACAGGGATCTGTCGAAGCAGGAGATTTATGATTCGTTTTCACATCGAAGCA 1758  
QY 5612 ATTAAGCTCAGGTTATGAGTAAACCGCGGATTTGTTTCGAAT 5659  
Db 1759 ATTAGACAGCTGGCTATAGCGTAATCACTCCAGTTGTCATTACGAAT 1806

## RESULT 13

US-09-221-017B-736/c  
; Sequence 736, Application US/09221017B

; Patent No. 6444799  
; GENERAL INFORMATION:

; APPLICANT: ROSS, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FASTSEQ for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221,017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Monroy, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-813-5600

; TELEFAX: 650-494-0792

; TELEX: 706141

; INFORMATION FOR SEQ ID NO: 736:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2295 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: UNKNOWN

; ORIGINAL SOURCE:

; ORGANISM: PORPHYROMONAS GINGIVALIS

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1...2295

; US-09-221-017B-736

Query Match 1.5%; Score 92; DB 4; Length 2295;

Best Local Similarity 51.6%; Pred. No. 4.1e-18;

Matches 285; Conservative 0; Mismatches 260; Indels 7; Gaps 3;

QY 1636 TGGGCTTGCACACAGATCCCTCACCACGTAGTACCTACCAAGAGCTCAATTCGCATGTATG 1695

Db 1642 TCGGTCTACCGACAGAGTTCTTCGCGATAGGAATGTACAGAGAGTAGTGAAGCATGCA 1583

QY 1696 AAGCTGGGAAAGTGTCAATTCAGAACTCAAGAGCATTTCTTGTGATGAATACGTGGGAC 1755

Db 1582 AAGAAGGATGCATATCTTTCCGCCACGTAATCACCTTCAACATGATGAATACGTAGGCC 1523

QY 1756 TAACCCGTGACATGAACACAGCTACTTTAAACCATTCGCAAGAGTTCACTGACCACA 1815

Db 1522 TTGCTATAGAGCATCCGAGAGCTATCATTCATTCATGCATCGGCATTTGTTCCACCATA 1463

QY 1816 TCGACATCGTTGATGAAGAGGTCTACAGCCAGATGTTGTCAAACCCCTGATC---CATACG 1872

Db 1462 TCGACATCTGCTCTCAGAACATCCACATTCCTCAATGGCAATGCTCTGACCTGACGGCG 1403

QY 1873 AAGCAGCTGCAGATATGAGCAAGATCGCTGAGAGATCCGTTGAGTTCAATCTCTG 1932

Db 1402 AGTGGATGCATACGAGAGAGCTATTGAAGCTCGCGGAGGATAGACCTTTTCATCGCG 1343

QY 1933 GCATCGCGGAAACGG-CACATCGCTTTTCATTGAACCATCATCTCTCTGTCAGGACTGA 1991

Db 1342 GGATAGGACCTGACGACATATCGCATTCATGAGCCGGGTCTCTCTTACCTCTCGCA 1283

QY 1992 CAAAGTCCAGCGCTGCACCTTAAACTGTGGAGGACACGCTCGATTCTTCAACACC- 2050

Db 1282 CCCGTATCAAGACGCTAACTACAGATACCCGTACTGGCTAACAGTCGCTTTTTCGACCA 1223

QY 2051 --ATCGAAGAGTCCCAACCCACGCGTCAACCGGTTTGGGACACTTTGTCGCGCGCG 2108

Db 1222 ATACGAATCAGGTTCCGAAAGTGTCTCACCGTAGGATAGGAACATATTATGATGCAC 1163

QY 2109 AAAACATCGTTGTTGGCACTGTGTAAGGAAAGCGACGCCATCCGCGGAATGTGG 2168

Db 1162 GGAAGTAATGATCTCTCGTGAACGACATACATAAAGCCCGTCTCTTCTGTAAGCTGTG 1103

QY 2169 AAGGCCAGTGA 2180

Db 1102 AAGTGCAGTCA 1091

## RESULT 14

US-09-367-293-1

; Sequence 1, Application US/09367293

; Patent No. 6444878

; GENERAL INFORMATION:

```

: APPLICANT: Donaldson, Lain A.
: APPLICANT: Bojsen, Kirsten
: APPLICANT: Jorgensen, Kirsten
: APPLICANT: Jorsboe, Morten
: TITLE OF INVENTION: SELECTION METHOD FOR TRANSGENIC PLANTS
: FILE REFERENCE: DY0021.001APC
: CURRENT APPLICATION NUMBER: US/09/367,293
: CURRENT FILING DATE: 1999-12-23
: PRIOR APPLICATION NUMBER: PCT/GB98/00367
: PRIOR FILING DATE: 1998-02-05
: PRIOR APPLICATION NUMBER: GB 9702592.8
: PRIOR FILING DATE: 1997-02-07
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 834
: TYPE: DNA
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (0)...(0)
: OTHER INFORMATION: Glucosamine-6-Phosphate Deaminase
: US-09-367-293-1

```

Query Match	1.5%	Score	90.6;	DB	4;	Length	834;
Best Local Similarity	48.7%;	Pred.	No.	5e-18;			
Matches	309;	Conservative	0;	Mismatches	319;	Indels	7;
	Gaps	2;					
QY	1636	TGGGGGTTCGAACAGAGTCTCTACCACTGAGTACCTACCAAGAGCTCATTCGCGATGTATG	1695				
Db	125	TGGGCTTGGGACTGGCGGACGCCGANTGCCACCTATTAAAGCGTTAGTCGAATGCATA	184				
QY	1696	AGCTTGGGAAGTGTCTATTCGAAGTATGCAAGGCATCTTTGTTGGATGAATACGTGGGAC	1755				
Db	185	AAGCAGGCCAGGTCAGCTTTAAGCAGCGTTGTCACTTCAACATGGAGCAATATGTGGTC	244				
QY	1756	TACCCGTGACGATGAACACAGCTACTTTTAAACCATTTCGAAAGAGTTTCACATGCACACA	1815				
Db	245	TGCGGAAGAGCATCGGAAGCTACTACAGCTTTATGACCCGTATTTCTTCGATCAAG	304				
QY	1816	TCGACATCGTTGATGAAGAGGTCTACAGCCGAGATGGTGCAAACCTGTATCCATACGAAG	1875				
Db	305	TTGATATTCCAGCAGAAACATCAAGCTTCTCAACGCAACGCCCGGATATCGAGCGCG	364				
QY	1876	CAGCTGCAGAGTATGAGGCAAGATCGCTGCAGATCCCGTTGAAGTTCAAATCCTTGGCA	1935				
Db	365	AGTGGCGCCAGTATGAAGAAANAATCGTTCTTACGGAAAAATTCATCTGTTTATGGCGG	424				
QY	1936	TCGGCGGAAAC----GSCACATCGCTTTCATTGAACCATCATCTCTCTCTCAGGACTGA	1991				
Db	425	GTGTAGTAAACGACGGTCATATTGCATTTAACGAACGGCGCTCTCTCTGGCTTCGTA	484				
QY	1992	CAGAAGTCCAGGGCGTCACCCCTAAACTGTGGAGNACACGCTCGATCTT---CAACA	2048				
Db	485	CTCGATATCAAAACCCCTGACTCATGACACTCGCGTCGCAAACTCTCGTTTCTTTGATAACG	544				
QY	2049	CCATCGAAGAGGTCCCAACCCAGCGCGCTACCCAGGGTTTGGGCACTTTGTCGCGCGCGC	2108				
Db	545	ATGTTAATCAGGTGCCAAATATGCCCTGACTGTCGGTGTGTGTACACTCTGGATGCCG	604				
QY	2109	AAACATCTGTTTGGTGGCAACTGTGTGAAGGAAGAACGCCGACGCCATCCGCGGAACCTGTGG	2168				
Db	605	AAGAAGTGAATTTCTGGTGTGGTAGCCAGAAAGCACTGGGGCTCGAGCGCGCGCGTTG	664				
QY	2169	AAGGCCAGTGACTGCTTTCTTGCCAGGTTCCATCCCTGTAGATGCACAAACATGCCACCAT	2228				
Db	565	AGGTTCCGTGAACCATATGTGGACCATCAGCTGTCTGCAACTGCCAAGCGATCA	724				
QY	2229	CATCGTTGGATGAAGCAGCAGTATCCAAAGCTGAA	2263				
Db	725	TGTTGTGCGATGAACCTTCCACCATTGAGCTGAA	759				

RESULT 15  
 US-09-557-884-1/c  
 : Sequence 1, Application US/09557884  
 : Patent No. 6506581  
 : GENERAL INFORMATION:  
 : APPLICANT: Fleischmann et al.  
 : TITLE OF INVENTION: The Nucleotide sequence of  
 : the Haemophilus influenzae Rd Genome, Fragments  
 : thereof, and Uses Thereof  
 :  
 : NUMBER OF SEQUENCES: 1  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Human Genome Sciences, Inc.  
 : STREET: 9410 Key West Avenue  
 : CITY: Rockville  
 : STATE: MD  
 : COUNTRY: USA  
 : ZIP: 20850  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 3 1/2 inch diskette  
 : COMPUTER: Dell Pentium  
 : OPERATING SYSTEM: MS DOS v6.22  
 : SOFTWARE: ASCII Text  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/557,884  
 : FILING DATE: 25-Apr-2000  
 : CLASSIFICATION: <Unknown>  
 :  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 08/476,102  
 : FILING DATE: JUN-5-1995  
 :  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Michelle S. Marks  
 : REGISTRATION NUMBER: 41,971  
 : REFERENCE/DOCLET NUMBER: PB186P3  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 301-309-8504  
 : TELEFAX: 301-309-8439  
 :  
 : INFORMATION FOR SEQ ID NO: 1:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1830121 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: linear  
 :  
 : SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1

us-10-019-284a-1.rni

Search completed: October 1, 2003, 01:26:45  
Job time : 282 secs

Search completed: October 1, 2003, 01:26:45

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 23:16:34 ; Search time 939 seconds

(without alignments)  
15825.103 Million cell updates/sec

Title: US-10-019-284A-1

Perfect score: 5969  
Sequence: 1 agtcctgcagccaccatt.....tgatatttgcgcgctgaa 5969

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1678620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	5656.2	94.8	3309400	10	US-09-738-626-1 Sequence 1, Appli
2	1874.2	31.4	12983	10	US-09-738-626-2904 Sequence 2904, Ap
3	1267.8	21.2	1299	10	US-09-738-626-2905 Sequence 2905, Ap
4	1110.4	18.6	1152	10	US-09-738-626-2907 Sequence 2907, Ap
5	671.8	11.3	759	10	US-09-738-626-2906 Sequence 2906, Ap
C 6	215.2	3.6	1059	11	US-09-746-660A-31 Sequence 31, Appl
7	192	3.2	975	10	US-09-974-300-829 Sequence 829, App
8	177.6	3.0	30246	8	US-08-781-986A-56 Sequence 56, Appl
9	174.2	2.9	1380	10	US-09-974-300-685 Sequence 685, App
10	163.6	2.7	1884	9	US-09-815-242-9369 Sequence 9369, Ap
11	161	2.7	1380	10	US-09-738-626-2903 Sequence 2903, Ap
12	153	2.6	5840	10	US-09-070-927A-31 Sequence 31, Appl
13	122.2	2.0	567	10	US-09-070-927A-31 Sequence 664, App
14	121.2	2.0	1143	14	US-09-070-927A-31 Sequence 925, Ap
15	121.2	2.0	9025608	14	US-10-156-761-1 Sequence 1, Appli
C 16	118.4	2.0	936	10	US-09-738-626-2908 Sequence 2908, Ap

17	117.4	2.0	783	14	US-10-156-761-3005 Sequence 3005, Ap
18	116	1.9	1098	10	US-09-974-300-734 Sequence 734, App
19	109.8	1.8	1014	10	US-09-974-300-822 Sequence 822, App
20	98.2	1.6	24004	10	US-09-070-927A-72 Sequence 72, Appl
21	97.2	1.6	882	10	US-09-974-300-5255 Sequence 5255, Ap
22	96.4	1.6	489	10	US-09-974-300-857 Sequence 857, App
C 23	90	1.5	1830121	14	US-10-329-960-1 Sequence 1, Appli
24	88.8	1.5	722	10	US-09-974-300-814 Sequence 814, App
C 25	86.4	1.4	7156	10	US-09-070-927A-29 Sequence 29, Appl
26	82.4	1.4	348	10	US-09-974-300-5313 Sequence 5313, Ap
27	79	1.3	480	11	US-09-918-995-21116 Sequence 21116, A
28	77.4	1.3	444	10	US-09-974-300-877 Sequence 877, App
29	77.2	1.3	474	10	US-09-070-927A-639 Sequence 639, App
30	77	1.3	1353	10	US-09-974-300-799 Sequence 799, App
31	76.8	1.3	1368	10	US-09-974-300-893 Sequence 893, App
32	75.6	1.3	1347	9	US-09-741-665-137 Sequence 137, App
33	75.6	1.3	1947	9	US-09-815-242-5992 Sequence 5992, App
C 34	74.2	1.2	9025608	14	US-10-156-761-1 Sequence 1, Appli
35	73.2	1.2	465	8	US-08-781-986A-1317 Sequence 1317, Ap
36	73	1.2	1398	14	US-10-156-761-3951 Sequence 3951, Ap
37	72.2	1.2	2130	9	US-09-079-892-6 Sequence 6, Appli
C 38	71.8	1.2	1216	14	US-10-106-698-647 Sequence 647, App
39	71.6	1.2	722	10	US-09-974-300-802 Sequence 802, App
C 40	71.2	1.2	9797	10	US-09-070-927A-550 Sequence 550, App
C 41	69.6	1.2	390	8	US-08-781-986A-1958 Sequence 1958, Ap
42	68.2	1.1	405	10	US-09-974-300-5262 Sequence 5262, Ap
43	67.8	1.1	429	10	US-09-974-300-5182 Sequence 5182, Ap
44	67.6	1.1	612	10	US-09-974-300-849 Sequence 849, App
45	64.4	1.1	1314	9	US-09-939-980-153 Sequence 153, App

#### ALIGNMENTS

RESULT 1

US-09-738-626-1/C

; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 94.8%; Score 5656.2; DB 10; Length 3309400;  
Best Local Similarity 97.1%; Pred. No. 0;  
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

QY 1 AGTCGTCGAGCCACCATTGATGCTGCTACCGAGCTTCGAGGCTTTTACATCT 60  
|||||

Db	2815585	AGTCGTCGACGCCACCATGATGAGGTGGTCAACAGAGCTTGGGAGGCTTTCTACATCC	2815526
Qy	61	ACGCTCCGTCGCGTGGAGTGGGTCATTAACGGTGGGATCACGGCGGTGAAAGTGGC	120
Db	2815525	ACACTGCGTCGCGTGGAGTGGGTCATTAACGGTGGGATCACGGCGGTGAAAGTGGC	2815466
Qy	121	GAACCCATGATGTTCCCTTGTGGGTTCAGGAACGAGTGGGTTGAAAGTTCGAAGNG	180
Db	2815465	GAACCCATGATGTTCCCTTGTGGGTTCAGGAACGAGTGGGTTGAAAGTTCGAAGNG	2815406
Qy	181	TCGTCAGTTTAAAGTTATGATCATCATCAGCTTGGAAAGGTGAGTTAAATCAGTAGACCTG	240
Db	2815405	TCGTCAGTTTAAAGTTATGATCATCATCAGCTTGGAAAGGTGAGTTAAATCAGTAGACCTG	2815346
Qy	241	CACACAGCGCTCAAGTCGAGAGATAATTAACCTAGATCCGAGACATAAGACATCATTA	300
Db	2815345	CACACAGCGCTCAAGTCGAGAGATAATTAACCTAGATCCGAGACATAAGACATCATTA	2815286
Qy	301	CGTCTATGCTTCTGCTGGAAGAACCAATAAATCACTCAGAAAGATGGCAGAGTGGTGCAAT	360
Db	2815285	CGTCTATGCTTCTGCTGGAAGAACCAATAAATCACTCAGAAAGATGGCAGAGTGGTGCAAT	2815226
Qy	361	ATCAAGAAATGAGGTCAGAGTCAAGAGTTAAATAAATGAGGGAAGAAATGTTCCCGCTTCG	420
Db	2815225	ATCAAGAAATGAGGTCAGAGTCAAGAGTTAAATAAATGAGGGAAGAAATGTTCCCGCTTCG	2815166
Qy	421	GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATACGGAACCTCTCTGGAGAAC	480
Db	2815165	GGGTGATGATGGCTTCTCCAACTCGAAACGGCATCATACGGAACCTCTCTGGAGAAC	2815106
Qy	481	CAGCACCTAAAGCGAGGATTCACCCCGAATCCACCGAGTGTCCCGGTTTATG	540
Db	2815105	CAGCACCTAAAGCGAGGATTCACCCCGAATCCACCGAGTGTCCCGGTTTATG	2815046
Qy	541	ATCTTCATATACAGGTGGAAACGGTGGCGCTTCTACGGGAACGACGACGCGGA	600
Db	2815045	ATCTTCATATACAGGTGGAAACGGTGGCGCTTCTACGGGAACGACGACGCGGA	2814986
Qy	601	GGACACCGCGCATATACCGGGAACATGCGCAGCAGCTGATGTTCCGCAAGCATGGTTT	660
Db	2814985	GGATGCGCGCAGTATACCGGGAACATGCGCAGCAGCTGATGTTCCGCAAGCATGGTTT	2814926
Qy	661	CGGCGCGGCTGACGCACTGCGCAGCGAGGTGGAACCTTATTCCTTGTGTGAAGAG	720
Db	2814925	CGGCGCGGCTGACGCACTGCGCAGCGAGGTGGAACCTTATTCCTTGTGTGAAGAG	2814866
Qy	721	TCCTGCTGTGGGCAATTCACCTCAGGGCCCTTTCATCAACGATGCGGTGTGGTGCTC	780
Db	2814865	GCCTGCTGTGGGCAATTCACCTCAGGGCCCTTTCATCAACGATGCGGTGTGGTGCTC	2814806
Qy	781	AAAACCGGATTTTCATTTTCCGGCAACCCACAGATCTTGCCCGGTGATCCATCGCG	840
Db	2814805	AAAACCGGATTTTCATTTTCCGGCAACCCACAGATCTTGCCCGGTGATCCATCGCG	2814746
Qy	841	GAAAGGTTGGATCAATTCGATCAGATAGCGCGGAACTGACAATCTTCTGAGGTTTC	900
Db	2814745	GAAAGGTTGGATCAATTCGATCAGATAGCGCGGAACTGACAATCTTCTGAGGTTTC	2814686
Qy	901	TCGATCTCTGGCAGCGCACCATCATCTGCTTCTTCCGGCACACTGATGCGAGATTTTG	960
Db	2814685	TCGATCTCTGGCAGCGCACCATCATCTGCTTCTTCCGGCACACTGATGCGAGATTTTG	2814626
Qy	961	ATACACCTACACGCGCAATTCGCTTGGCTAAAGAGAAAAATGTGACGGTCACGGCTACGC	1020
Db	2814625	ATACACCTACACGCGCAATTCGCTTGGCTAAAGAGAAAAATGTGACGGTCACGGCTACGC	2814566
Qy	1021	ATTGTTCAATGCGATGCTCCGCTGCATCATATAGGGCTCCGGCAGCGTGGCGCTTTGC	1080
Db	2814565	ATTGTTCAATGCGATGCTCCGCTGCATCATATAGGGCTCCGGCAGCGTGGCGCTTTGC	2814506
Qy	1081	TGCTGCGGCACTGCGGGGACCATATGTTGAGTTGATCGCGGACGGGTGCATTTGG	1140
Db	2814505	TGCTGCGGCACTGCGGGGACCATATGTTGAGTTGATCGCGGACGGGTGCATTTGG	2814446
Qy	1141	CCGATGGAACGGTGCATCTAGCTCGTCCCAACACGCTTTTTCATCACGAGCCCATGG	1200
Db	2814445	CCGATGGAACGGTGCATCTAGCTCGTCCCAACACGCTTTTTCATCACGAGCCCATGG	2814386
Qy	1201	AAGCCGCGGAATGCCACACGGTGTAGTACATTTTGGGCGTTTGAACGTACCGTCAACG	1260
Db	2814385	AAGCCGCGGAATGCCACACGGTGTAGTACATTTTGGGCGTTTGAACGTACCGTCAACG	2814326
Qy	1261	ATGAGATGCGCGCTCTGGCGATGGCGGCATTCGCGGGGACACGACGACATAGCGA	1320
Db	2814325	ATGAGATGCGCGCTCTGGCGATGGCGGCATTCGCGGGGACACGACGACATAGCGA	2814266
Qy	1321	GTCACTTGTGTCACACACGGTGTAGTACGCTTATCGAGCGGACCTCCACACCT	1380
Db	2814265	GTCACTTGTGTCACACACGGTGTAGTACGCTTATCGAGCGGACCTCCACACCT	2814206
Qy	1381	CAACCGTCCCGCTTAAATTTCTCGTCTTGGGATCAGAAATGCTTAAATCCACCTCG	1440
Db	2814205	CAACCGTCCCGCTTAAATTTCTCGGACATTCGAGATCAGAAATGCTTAAATCCACCTCG	2814146
Qy	1441	CAATTTTGTGTTTGTACTCAACGGCCAGGTGCAAAAGTCCATTTAGTGCATCAAG	1500
Db	2814145	TAATTTTGTGTTTGTACTCAACGGCCAGTTCACACAGTCCATTTAGACCATCAAG	2814086
Qy	1501	TACTTTAGTAGTAGTAAACTATCTCTGAT-TTTAAAGAGTCCCACCATGGAATCACT	1559
Db	2814085	TAATTTAAATACGACCAAACTTCTGATATAAAGAGTCCGACCATGGACATCATC	2814026
Qy	1560	ATCTGCAAAAGCAGCAGAAAGTTCGCAAGCAGTTGCACTTAAATCGACCCCTTCGCC	1619
Db	2814025	ATCTGCAAAAGCAGCAGAAAGTTCGCAAGCAGTTGCACTTAAATCGACCCCTTCGCC	2813966
Qy	1620	AACAAGGTGGAACTTGGGGTTCACACAGAGATCTCACCACTGAGTACCTACCAAG	1679
Db	2813965	ACTAAGGGCGGAACCTTGGGGTTCACCTGAGTCTCACCTTTCAGCAGCTACCAAG	2813906
Qy	1680	CTCATTCGATGTATGAAGCTGGGAAGTGTTCATTCAGAACTGCAAGGATCTTGTGTTG	1739
Db	2813905	CTCATTCGATGTATGAAGCTGGGAAGTGTTCATTCAGAACTGCAAGGATCTTGTGTTG	2813846
Qy	1740	GATGAATACGTGGGACTAACCCGTGACGATGAAACACAGCTACTTTAAACCATTCGCAAA	1799
Db	2813845	GATGAATACGTGGGATTAACCGCGCAGATGAAACACAGCTACTTTAAACCATTCGTA	2813786
Qy	1800	GAGTTCACGTGACCATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCAAC	1859
Db	2813785	GAGTTCACGTGACCATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCAAC	2813726
Qy	1860	CCTGATCCATACGACGCTGACAGTATGAGGCAAGATCGCTGCAAGATCCGTTGAA	1919
Db	2813725	CCTGATCCATACGACGCTGACAGTATGAGGCAAGATCGCTGCAAGATCCGTTGAT	2813666
Qy	1920	GTTCAATCCTTGGCATCGCGGAAACGG-CACATCGCTTTCATTTGAACCATCATCTCT	1978
Db	2813665	GTTCAATCCTTGGCATCGCGGAAACGG-CACATCGCTTTCATTTGAACCATCATCTCT	2813606
Qy	1979	CTGTCAGGACTGACAAAGTCCAGGCGCTGCAACCTTAAACCTGTGGAGGACCACTCGA	2038
Db	2813605	CTGTCAGGACTGACAAAGTCCAGGCGCTGCAACCTTAAACCTGTGGAGGACCACTCGA	2813546
Qy	2039	TTCCTCAACACCATCGAAGAGTCCCAACCGCGGTCCACCGAGGTTTGGGCACTTGG	2098
Db	2813545	TTCCTCAACACCATCGAAGAGTCCCAACCGCGGTCCCAACCGAGGTTTGGGCACTTGG	2813486
Qy	2099	TCCCGCGCAAAACATCTGTTGTTGGCAACTGTTGAAGGAAAGCCGACCCATCCGC	2158
Db	2813485	TCCCGCGCAAAACATCTGTTGTTGGCAACTGTTGAAGGAAAGCCGACCCATCCGC	2813426
Qy	2159	GGAACGTGTGAAGGCCCATGACTGCTTCCGACAGGTTTCCATCTGTAGATGCACAC	2218
Db	2813425	GGAACGTGTGAAGGCCCATGACTGCTTCCGACAGGTTTCCATCTGTAGATGCACAC	2813366

QY	2219	ATGCCACCATCATCGTTGGATGAAGCAGCAGTATCCAAGCTCGAAGAACGCTGATCACTAC	2278
Db	2813365	AATGCCACCATCATCGTTGATGAAGCAGCAGCATCCAAGCTCGAAGAACGCTGATCACTAC	2813306
QY	2279	CGTCTCATGGACCAATTAAGCTGGCTAGAACAAAGAAAGAACTACTGTGTGGGGCTA	2338
Db	2813305	CGTCTCATGGACCAATTAAGCTGCCTAGAACAAAGAAAGTGTGTGGGGCTA	2813246
QY	2339	TGCACACAGAACTTTCAGTTTGGCCCTTCGCTACCATGTGACTCCTCCGACGGCAGGC	2398
Db	2813245	TGCACACAGAACTTTCAGTTTGGCCCTTCGCTACCATGTGACTCCTCCGACGGCAGGC	2813186
QY	2399	TCAATGATCCAAAGGAATGTACGTGATGGAGATACCTCCACGTCTACTACGACGAGC	2458
Db	2813185	TCAATGATCCAAAGGAATGTACGTGATGGAGATACCTCCACGTCTACTACGACGAGC	2813126
QY	2459	ATCCAGGTTTCCCTTCGCGACCAAGCGACCGCTGGCTCAGACACGACGCGGTTGA	2518
Db	2813125	ATCCAGGTTTCCCTTCGCGACCAAGCGACCGCTGGCTCAGACACGACGCGGTTGA	2813066
QY	2519	CCGGACCGACGATTTGCAGTGGACGCACTGCCCGACGCTCTTTACCCGGATGATCCT	2578
Db	2813065	CCGGACCGACGATTTGCAGTGGACGCACTGCCCGACGCTCTTTACCCGGATGATCCT	2813006
QY	2579	ATCACTGGATGATGCTATTCCGGTGGAGCCGATTTACTGACGSCACACTAAACTTT	2638
Db	2813005	ATCACTGGATGATGCTATTCCGGTGGAGCCGATTTACTGACGSCACACTAAACTTT	2812946
QY	2639	TCTACACCGCAACCTTAAATTTAGCGAAGCGCCGCGACCCAAAACCTTGTGGAAG	2698
Db	2812945	TCTACACCGCAACCTTAAATTTAGCGAAGCGCCGCGACCCAAAACCTTGTGGAAG	2812886
QY	2699	TCGAGGACCAACTGGCTGATGGGGGATTCATCGCGTTCGCCCTAAAAATCCGCTTA	2758
Db	2812885	TCGAGGACCAACTGGCTGATGGGGGATTCATCGCGTTCGCCCTAAAAATCCGCTTA	2812826
QY	2759	TCGACGACCGCCAGCGGTTTCAACCCCATTTACCGCGATCCCATGATCAGCCCTGATG	2818
Db	2812825	TCGACGACCGCCAGCGGTTTCAACCCCATTTACCGCGATCCCATGATCAGCCCTGATG	2812766
QY	2819	GTGATGTTTGGAACTGGTTCTTGGGGCCCAACGCGAAACCTCACCGGTGAGCGGTTTC	2878
Db	2812765	GTGATGTTTGGAAATGGTTCTTGGGGCCCAACGCGAAACCTCACCGGTGAGCGGTTTC	2812706
QY	2879	TATACCGCTCGACAGATCTTTGAAACTGGGAATTTCCCGTGAAATCACTTTGACCTCA	2938
Db	2812705	TATACCGCTCGACAGATCTTTGAAACTGGGAATTTCCCGTGAAATCACTTTGACCTCA	2812646
QY	2939	GTGATGACAACTGGTTCTGCTCCGTGATCTCGTTCCCGATGGCTACATGGGGATGCC	2998
Db	2812645	GTGATGACAACTGGTTCTGCTCCGTGATCTCGTTCCCGATGGCTACATGGGGATGCC	2812586
QY	2999	CCAACCTTTTACGCTTCGGCGATGAAGAACTGGCGAAGATCTCGAGTGCCTGATTTCT	3058
Db	2812585	CCAACCTTTTACGCTTCGGCGATGAAGAACTGGCGAAGATCTCGAGTGCCTGATTTCT	2812526
QY	3059	GTCCACAAGGATTTGGCCGAATCCACGATGAGTTTACTACTACGCAAGCTCTGACAGT	3118
Db	2812525	GTCCACAAGGATTTGGCCGAATCCACGATGAGTTTACTACTACGCAAGCTCTGACAGT	2812466
QY	3119	GCGGATATGTCGTGCAAGCTTGAAGGAACGACCTTCCCGCTCTTCGCGAGGATTCAGCG	3178
Db	2812465	GCGGATATGTCGTGCGCAAGCTTGAAGGAACGACCTTCCCGCTCTTCGCGAGGATTCAGCG	2812406
QY	3179	AGCTGGATTTGGGCCATGAATTTACGCAACCGAGGTTGCAGTAACGGTTTCTGATGCCT	3238
Db	2812405	AGCTGGATTTGGGCCATGAATTTACGCAACCGAGGTTGCAGTAACGGTTTCTGATGCCT	2812346
QY	3239	GGCTCGTGGGTGATGGGGTCCCGCGCAGGATGATCAACCAACAGTTTGCACGAAG	3298
Db	2812345	GGCTCGTGGGTGATGGGGTCCCGCGCAGGATGATCAACCAACAGTTTGCACGGGAAG	2812286
QY	3299	GATGGGTGCATGCTGACTGTGCCCGCAAGCTTCATTTGGCGCAACCGACGATCTACC	3358

Db	2812285	GATGGGTGCACTGCTCTGACATGTCGCCCGAAGCTTCATTTGCGAACCCAGCATCTATC	28122222
Qy	3359	AAGAGCTCCCTTCTCCAGAGGGGAGTCGGGGTAATCAGATCTGTATTAGGTCTCGAAC	3418
Db	2812225	AAGAGCTTCTTCTCCAGAGGGGAGTCAGGGGTAAATCAGATCTGTATTAGGTCTCGAAC	2812166
Qy	3419	CTGTCCAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGCTGTCGGTTTGTCTG	3478
Db	2812165	CTGTCCAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGCTGTCGGTTTGTCTG	2812106
Qy	3479	TGGATCGTATGTTGATCGTCGCGTAGCTCAGGTAAACCTGGCGAATTAGTGATCGCGG	3538
Db	2812105	TGATCTGGTGGTGATCGTCGCGTAGCTCAGGTAAACCTGGCGAATTAGTGATCGCGG	2812046
Qy	3539	ACGATAATACAGCCATTGAGATAACTCCAGGTGATGGACAGGTTTCATTCGCTTTTCCGG	3598
Db	2812045	ACGATAATACAGCCATTGAGATAACTCCAGGTGATGGACAGGTTTCATTCGCTTTTCCGG	2811987
Qy	3599	GCCTTCAAAGGTGACACTATTGAGAGATAAGTCATATAAAGGTCTTTTGTGGCGAATT	3658
Db	2811986	GCCTTCAAAGGTGACACTATTGAGAGATAAGTCATATAAAGGTCTTTTGTGGCGAATT	2811927
Qy	3659	GTACAAATACTTCGCAAAATCCCTTGATCGGACACAAATAACAGGTTTAAATTTGTTTA	3718
Db	2811926	GTACAAATACTTCGCAAAATCCCTTGATCGGACACAAATAACAGGTTTAAATTTGTTTA	2811867
Qy	3719	GCCTTTGAACAAACATTCACTGCTCGAATATTTTGTTCCTCCGGTTAAGAGAGAAATTC	3778
Db	2811866	GCCTTTGAACAAACATTCACTGCTCGAATATTTTGTTCCTCCGGTTAAGAGAGAAATTC	2811807
Qy	3779	ATGACCATAAGGACCTCGCGCAACGCATCCTCGCGACATTTGCGGCGAGACACAAT	3838
Db	2811806	ATGACCATAAGGACCTCGCGCAACGCATCCTCGCGACATTTGCGGCGAGACACAAT	2811747
Qy	3839	GTCCGCGCGCACACTGTGCACACCGGTTTAGCCCTCGCTCRAAGACACCAAGGATGTG	3898
Db	2811746	GTCCGCGCGCACACTGTGCACACCGGTTTAGCCCTCGCTCRAAGACACCAAGGATGTG	2811687
Qy	3899	GATGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCACTTTGAACTGGCGGCATG	3958
Db	2811686	GATGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCACTTTGAAACGGTGGTATG	2811627
Qy	3959	TTCAGATCATCTGCGGCCAGGCGATGTGGATCATGTTTCAAGAACTCGATGACGCA	4018
Db	2811626	TTCAGATCATCTGCGGCCAGGCGATGTGGATCATGTTTCAAGAACTCGATGACGCA	2811567
Qy	4019	ACCTCCAAGACATCGCTGTCTCCACAGAGCAGCTCAAAGATGTTGTGCTTAACAAGCC	4078
Db	2811566	ACCTCCAAGACATCGCTGTCTCCACAGAGCAGCTCAAAGATGTTGTGCTTAACAAGCC	2811507
Qy	4079	AACGTGTTACGCGTCTGTGAAGTATTGGCGGACATTTTCGTCCCGCTGATTCCTATC	4138
Db	2811506	AACGTGTTACGCGTCTGTGAAGTATTGGCGGACATTTTCGTCCCGCTGATTCCTATC	2811447
Qy	4139	TGTTTGTGTGCGGTCCTCATGTGCTATCAACATGTTGGTTGCGCAGGATCTGTTC	4198
Db	2811446	TGTTTGTGTGCGGTCCTCATGTGCTATCAACATGTTGGTTGCGCAGGATCTGTTC	281138
Qy	4199	GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTCGTAGATGATCAAC	4258
Db	2811386	GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTCGTAGATGATCAAC	2811322
Qy	4259	CTGATGGCATCTGCCCGCTTCGCGTCTTTCACAGTGTGGTTTTCACCGCAACCAAG	4318
Db	2811326	CTCATGGCATCTGCCCGCTTCGCGTCTTTCACAGTGTGGTTTTCACCGCAACCAAG	2811266
Qy	4319	CGTTTCGGTGGCAATGAGTTCCTGGGCGCGGCATTTGTTATGGCGATGGTGTCCCAACC	4378
Db	2811266	CGTTTCGGGCAATGAGTTCCTGGGCGCGGCATTTGTTATGGCGATGGTGTCCCGGAGC	2811207
Qy	4379	CTGGTTACGGGTACGAGTGGCCGCCACCATGACCGCGGGGCAATGCCAATGTGGTCC	4438

```
Db 2811206 TTGGTGAACGGCTACGACGTGGCCGCCACCATGGCTGCGGGCGAAATGCCAATGTGGTCC 2811147
QY 4439 CTGTTGGTTTGGATGTTGCTCAAGCTGTTTACAGGCAACGCTGCTTCCTGTGCTGGT 4498
Db 2811146 CTGTTGGTTTGGATGTTTCCCAAGCCGGTTTACCAGGCGACCGTCTTCCTGTGCTGGT 2811087
QY 4499 GTCTCTTGATCTGCAACGATCGAGAGTTCTTGCACACGACTCATGGGCACTGCA 4558
Db 2811086 GTTCTTGATCTGCAACGATCGAGAGTTCTTGCACACGACTCATGGGCACTGCA 2811027
QY 4559 GACTTCTGATCACCCCAAGTGTGACTCTGTGCTCACCGGCTTCTTACGTTTCACTT 4618
Db 2811026 GACTTCTGATCACCTCCAGTGTGACTCTGTGCTCACCGGATTCCTTACATTCATGCC 2810967
QY 4619 ATTGTCACGACATCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4678
Db 2810966 ATTGGCCCAAGCATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2810907
QY 4679 GATTTCGGTGGTCCAGTCCGCTGCTGCTTTTCGGTCTGCTGCTTACTCACCATTCTTATC 4738
Db 2810906 GATTTCGGTGGTCCAGTCCGCTGCTGCTTTTCGGTCTGCTGCTTACTCACCATTCTTATC 2810847
QY 4739 ACTGGTCTGCACGACGCTTCCCGCCCAATTGAGCTGGAGCTGTTCAACGAGGTGGATCC 4798
Db 2810846 ACTGGTCTGCACGACGCTTCCCGCCCAATTGAGCTGGAGCTGTTTAAACGAGGTGGATCC 2810787
QY 4799 TTCACTCTGCAACCCGATCCATGCGCAATATCGCGCAGGTCGACGATGTTTGGCAGTG 4858
Db 2810786 TTCACTCTGCAACGACGCTTATGGCTATATCGCCACGGTGGCGCATGTTTGGCAGTG 2810727
QY 4859 TTTCTTCTAGCAAGAGTGAAGAAGTCAAGGCGCTTGCAGGTGCTTCAGGTCTCCCGT 4918
Db 2810726 TTTCTTCTGCGCAAGAGTGAAGAAGTCAAGGCGCTTGCAGGTGCTTCAGGTCTCCCGT 2810667
QY 4919 GTTCTGGTATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGGCTGGCTGGCGGTC 4978
Db 2810666 GTTCTGGTATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGGCTGGCTGGCGGTC 2810607
QY 4979 TACATGTTATGCTGACGAGCTATCGGTGGCGCTTGAATGCACTCTTTGATATCAAG 5038
Db 2810606 TTCATCGGTATCGGTATCCGACGCTATCGGTGGCGCTTGAATGCACTCTTTAATATCAAG 2810547
QY 5039 GCAGTTGCGTGGCGGCTGAGGTTCTTCTGGTGTGTTTCTTATGATGCTCCAGATAG 5098
Db 2810546 GCAGTTGCGTGGCGGCTGAGGTTCTTCTGGTGTGTTTCTTATGATGCTCCAGATAG 2810487
QY 5099 GTCATGTTCTGTTGGCGGTTAGTACCTTTGTCATCGCATTCGGCGCAGCATGCT 5158
Db 2810486 GTCATGTTCTGTTGGTGTGAGTGTGTACCTTTCTCATCGCATTCGGCGCAGCATGCT 2810427
QY 5159 TATGGCTTTACTTGTGTCGCCAACGCGCAGCATTTGATCCAGATCCACCGCTGCTCCA 5218
Db 2810426 TATGGCTTTACTTGTGTCGCCAACGCGCAGCATTTGATCCAGATCCACCGCTGCTCCA 2810367
QY 5219 GTCCGTGCAAGAACGCAACGCAAGCAGAGCAGCCGCAAGATTTTCAACAGATTC 5278
Db 2810366 GTCCGTGCAAGAACGCAACGCAAGCAGAGCAGCCGCAAGATTTTCAACAGATTC 2810307
QY 5279 ACCATCATCAGCAGCTTTGACCGGTGAAGCTATCGCACTGAGCAGCTCAGCGATGCC 5338
Db 2810306 ACCATCATCAGCAGCTTTGACCGGTGAAGCTATTTGCATGAGCAGCTCAGCGATGCC 2810247
QY 5339 ATGTTTCCACGCGAAAGCTTGGCTCAGGTGTTGCGATCGTCCGCCAACGAGGCGAGTG 5398
Db 2810246 ATGTTTCCACGCGAAAGCTTGGCTCAGGTGTTGCGATCGTCCGCCAACGAGGCGAGTG 2810187
QY 5399 GTTTCACCATGAGCGGAAGATCGTGGTGGCTTCCCATCTGGTCAACGCTTTCGAGTC 5458
Db 2810186 GTTTCCTCGGTGAGTGAAAGATGTGGTGGCATTCCTCATCTGGCCATGCTTTCGAGTT 2810127
QY 5459 CGCACTAAGGCTGAGGATGTTTCCAAATGTGGATATCTTGTATGACATGTTGTTTCGACAC 5518
Db 2810126 CGCACTAAGGCTGAGGATGTTTCCAAATGTGGATATCTTGTATGACATGTTGTTTCGACACA 2810067
QY 5519 GTAAACCTCAACGCGCAGCACTTTAACCCGCTGAAGAAGCAGGCGGATGAAGTCAAGCA 5578
Db 2810066 GTAAACCTCAACGCGCAGCACTTTAACCCGCTGAAGAAGCAGGCGGATGAAGTCAAGCA 2810007
QY 5579 GGGGAGCTGCTGTGTGAATTCGATTAATGATCCATTAAGCTGCGAGGTTATGAGGTAACC 5638
Db 2810006 GGGGAGCTGCTGTGTGAATTCGATTAATGATCCATTAAGCTGCGAGGTTATGAGGTAACC 2809947
QY 5639 ACGCCGATGTTGTTTTCGAATTACAAGAAACCGGACCTGTAAACACTTACGGTTTGGGC 5698
Db 2809946 ACGCCGATGTTGTTTTCGAATTACAAGAAACCGGACCTGTAAACACTTACGGTTTGGGC 2809887
QY 5699 GAAATTGAAGCGGGGAGGCAACCTGCTCAACGTCGCAAGAAAGAGGGTCCAGCAACA 5758
Db 2809886 GAAATTGAAGCGGGGAGGCAACCTGCTCAACGTCGCAAGAAAGAGGGTCCAGCAACA 2809827
QY 5759 CCATAAGTTGAACCTTGTGAGTGTTCGCACACAGGTTAGATAGGGGACGTGACTACGC 5818
Db 2809826 CCATAAGTTGAACCTTGTGAGTGTTCGCACACAGGTTAGATAGGGGACGTGACTACGC 2809767
QY 5819 ATCTTTGACACCGGTACCGGTACCGTTCGAGATTTTAAACCTGTTCAACGAGTCAAGC 5878
Db 2809766 ATCTTTGACACCGGTACCGGTACCGTTCGAGATTTTAAACCTGTTCAACGAGTCAAGC 2809707
QY 5879 TCGGTGTACCTGTGTGTGTCGCCACCCGCAATCTTACCCACATTTGGACATTTGTCGTCA 5938
Db 2809706 TCGGTGTACCTGTGTGTGTCGCCACCCGCAATCTTACCCACATTTGGACATTTGTCGTCA 2809647
QY 5939 GCAGTAGCGTTTGATATTTTGGCGCGGTG 5967
Db 2809646 GCAGTAGCGTTTGATATTTTGGCGCGGTG 2809618

RESULT 2
US-09-738-626-2904
; Sequence 2904, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2904
; LENGTH: 1983
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2904

Query Match 31.4%; Score 1874.2; DB 10; Length 1983;
Best Local Similarity 96.6%; Pred.No. 0;
Matches 1915; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 3779 ATGGACCAATAGGACCTCGCAAGCATCTCTGCGCAGATTTGGCGGCAAGACAACATT 3838
|||||
```



APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: JP 00/280988  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 2905  
LENGTH: 1299  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-2905

Query Match 21.2%; Score 1267.8; DB 10; Length 1299;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1286; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 2327 TGTGTGGGCTATGCACAGAACTTCCAGTTTGGCCCTCGCTACCATGTGACTCCTC 2386  
DB 2 TGTGTGGGCTATGCACAGAACTTCCAGTTTGGCCCTCGCTACCATGTGACTCCTC 61  
QY 2387 CGCAGGCGAGGCTCAATGATCCCAACGGAATGATCGATGAGATACCCCTCCAGTCT 2446  
DB 52 CGCAGGCGAGGCTCAATGATCCCAACGGAATGATCGATGAGATACCCCTCCAGTCT 121  
QY 2447 ACTACAGCAGATCCAGGTTTCCCTTCGACCAAGCGCCGCTGGGCTCACACCA 2506  
DB 122 ACTACAGCAGATCCAGGTTTCCCTTCGACCAAGCGCCGCTGGGCTCACACCA 181  
QY 2507 CCACGCGGTTGACCGGACCGCAGATGCGAGTGCAGCAGCAGCTGCCGAGCTTTTACC 2566  
DB 182 CCACGCGGTTGACCGGACCGCAGATGCGAGTGCAGCAGCAGCTGCCGAGCTTTTACC 241  
QY 2567 CGGATGATCCTATGACCTGGATGATGCTATTCGGTGGAGCGCTATTACTGAGCGCA 2626  
DB 242 CGGATGATCCTATGACCTGGATGATGCTATTCGGTGGAGCGCTATTACTGAGCGCA 301  
QY 2627 CACTTAACTTTTACACCGGCAACCTAAATTTGACGAAAGCGCCGCGCCACCAAA 2686  
DB 302 CACTTAACTTTTACACCGGCAACCTAAATTTGACGAAAGCGCCGCGCCACCAAA 361  
QY 2687 ACCTTGTGAGTCGAGGACCGCACTGGGCTGATGGGCGGCATTCATCGCGTTCCGCTA 2746  
DB 362 ACCTGTCGAAGTCGAGGACCGCACTGGGCTGATGGGCGGCATTCATCGCGTTCCGCTA 421  
QY 2747 AAAATCCGCTTATCGAGGACCGCGGCTTTCACACCCCATTTACCGGATCCCATGA 2806  
DB 422 AAAATCCGCTTATCGAGGACCGCGGCTTTCACACCCCATTTACCGGATCCCATGA 481  
QY 2807 TCAGCCCTGATGTGATGTTGGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCG 2866  
DB 482 TCAGCCCTGATGTGATGTTGGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCG 541  
QY 2867 GTGAGGCGGTTCTATACCGCTGCACAGATCTTGAAGAACTGGGAATTCCTCGGTGAATCA 2926  
DB 542 GTGAGGCGGTTCTATACCGCTGCACAGATCTTGAAGAACTGGGAATTCCTCGGTGAATCA 601  
QY 2927 CCTTTGACCTCAGTATGACACACCTGGTTCTGCTGATCTGCTTCCCGATGGGTACA 2986  
DB 602 CCTTTGACCTCAGTATGACACACCTGGTTCTGCTGATCTGCTTCCCGATGGGTACA 661  
QY 2987 TGTGGGAATCCCAACCTTTTACGCTTCGCGATGAAGAACTGGGAGATCTCGAGC 3046  
DB 562 TGTGGGAATCCCAACCTTTTACGCTTCGCGATGAAGAACTGGGAGATCTCGAGC 721

QY 3047 TGCTGATTTTCTGTCCACAGGATTGGACCGAATCCACGATGAGGTTACTCAGTACGCAA 3106  
DB 722 TGCTGATTTTCTGTCCACAGGATTGGACCGAATCCACGATGAGGTTACTCAGTACGCAA 781  
QY 3107 GCTCTGACCACTGGCGGATATGCTGCGACAAGCTTGAAGAACACCTTCGCGCTCTTGC 3166  
DB 782 GCTCTGACCACTGGCGGATATGCTGCGACAAGCTTGAAGAACACCTTCGCGCTCTTGC 841  
QY 3167 GAGGATTCAGGAGAGCTGGATTTCGGCCCATGAATTCACGACCCGAGGTTGACGATAAAG 3226  
DB 842 GAGGATTCAGGAGAGCTGGATTTCGGCCCATGAATTCAGCACCAGGTTGACGATAAAG 901  
QY 3227 GTTCTGATGCTTGGCTCTGCTGGGATGGGCTGCCCGCCAGGATGATCACCACAG 3286  
DB 902 GTTCTGATGCTTGGCTCTGCTGGGATGGGCTGCCCGCCAGGATGATCACCACAG 961  
QY 3287 TTGCACAGGAGGATGGGTGCACCTGCTGCTGCTGCTGCCCGCAAGCTTCATTTCGCGCAAC 3346  
DB 962 TTGCACAGGAGGATGGGTGCACCTGCTGCTGCTGCCCGCAAGCTTCATTTCGCGCAAC 1021  
QY 3347 AGCGATCTACAAAGAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAATCAGATCTGTAT 3406  
DB 1022 AGCGATCTACAAAGAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAATCAGATCTGTAT 1081  
QY 3407 TAGGTTCTGAACCTGCTCGAGTAGACATCCGAGGCAATATTTCCTCGAGTGGGATGGT 3466  
DB 1082 TAGGTTCTGAACCTGCTCGAGTAGACATCCGAGGCAATATTTCCTCGAGTGGGATGGT 1141  
QY 3467 TCCGTTTCTGTGTGATCGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3526  
DB 1142 TCCGTTTCTGTGTGATCGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1201  
QY 3527 TAGTGTATCGGAGCAGTAATACAGCATTGAGTAACCTGAGGTTGATGAGGAGGTTTCAT 3586  
DB 1202 TAGTGTATCGGAGCAGTAATACAGCATTGAGTAACCTGAGGTTGATGAGGAGGTTTCAT 1261  
QY 3587 TCGCTTTTCCGGGCTTCAAAGGTGACACTATTGAGAGA 3625  
DB 1262 TCGC-TTTCGGGCTTCAAAGGTGACACTATTGAGAGA 1299

## RESULT 4

US-09-738-626-2907  
; Sequence 2907, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 2907  
; LENGTH: 1152  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum

US-09-738-626-2907

Query Match 18.6%; Score 1110.4; DB 10; Length 1152;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 1126; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 354 GTGCATTATCAAGAAATGAGGTCAAGCAGTTAAATAAATTTAGGGAAGAATTTGTTC 413  
DB 1 GTGCATTATCAAGAAATGAGGTCAAGCAGTTAAATAAATTTAGGGAAGAATTTGTTC 60

QY 414 CCCCTCGGGGTGATTGATGGCTTTCTCAACTCGAAGACGGCATCATCAGGAACTCTCT 473  
DB 61 CCCACGGGTGATTGATGGCTTTCTCAACTCGAAGACGGCATCATCAGGAACTCTCT 120

QY 474 GGAGAACAGCAGCTAAAGCGCAGGATTCACCCCGAATCCCGACGATTTGTCGGGT 533  
DB 121 GGAGAACAGCAGCTAAAGCGCAGGATTCACCCCGAATCCCGACGATTTGTCGGGT 180

QY 534 TTTATTGATCTTCAATACAGGTGGAAACGGTGGCGGTTTCTACGGGAACGAGGAC 593  
DB 181 TTTATTGATCTTCAATACAGGTGGAAACGGTGGCGGTTTCTACGGGAACGAGGAC 240

QY 594 CAGCGGAACACCGGCAGTATCACCGCAACATGCGACGACCGGTGATTTCCCAAGC 653  
DB 241 CAGCGGAACACCGGCAGTATCACCGCAACATGCGACGACCGGTGATTTCCCAAGC 300

QY 654 ATGTTTCGGCGCGGTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTTGTGT 713  
DB 301 ATGTTTCGGCGCGGTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTTGTGT 360

QY 714 GAAGAGTCTCTGTCGGGATTCACCTCAGGGCCGTTTCATCAACGATCGCGTTGT 773  
DB 361 GAAGAGTCTCTGTCGGGATTCACCTCAGGGCCGTTTCATCAACGATCGCGTTGT 420

QY 774 GTGCTCAAAACCGGATTTTATTTCCCGCAACCAACAGATCTTTCGCCGGTGATC 833  
DB 421 GTGCTCAAAACCGGATTTTATTTCCCGCAACCAACAGATCTTTCGCCGGTGATC 480

QY 834 CATCGGGAAGAAGTTGATCAATCGATCAGTAGCGCGGGAACATGACAACTTTCT 893  
DB 481 CATCGGGAAGAAGTTGATCAATCGATCAGTAGCGCGGGAACATGACAACTTTCT 540

QY 894 GAGCTTCGATCTCTCGGAGCGACCAACATCATTTGCTTCTCGGCACTGATGCA 953  
DB 541 GAGCTTCGATCTCTCGGAGCGACCAACATCATTTGCTTCTCGGCACTGATGCA 600

QY 954 GATTTGTATACACTACAGCGCAATTCCTTGGCTGAAGAAATGTGACGGTACG 1013  
DB 601 GATTTGTATACACTACAGCGCAATTCCTTGGCTGAAGAAATGTGACGGTACG 660

QY 1014 GCTAGCATTTGTTCAATGCGATGCTCGGTCATCATAGGCTCCCGCAGGTGGC 1073  
DB 661 GCTAGCATTTGTTCAATGCGATGCTCGGTCATCATAGGATCCCGCAGGTGGC 720

QY 1074 GCTTTGCTGTCGCGCAGTCCCGGGACGATATGTTGATTCGCGGCGCGGTG 1133  
DB 721 GCTTTGCTGTCGCGCAGTCCCGGGACGATATGTTGATTCGCGGCGCGGTG 780

QY 1134 CATTTGGCGGATGGAACGGTCTAGCTCTGTTCCACACGCGCTTTTTCATCAGGAC 1193  
DB 781 CATTTGGCGGATGGAACGGTCTAGCTCTGTTCCACACGCGCTTTTTCATCAGGAC 840

QY 1194 GCATGGGAAGCGCGGATGCAACGGTGTGATTTGCGGCTTTTGAACGTGACC 1253  
DB 841 GCATGGGAAGCGCGGATGCAACGGTGTGATTTGCGGCTTTTGAACGTGACC 900

QY 1254 GTCACGGATGAGTCCCGCTGTCGGCATGCGCGGCGCATCGCGGGGGACACAGACA 1313  
DB 901 GTCACGGATGAGTCCCGCTGTCGGCATGCGCGGCGCATCGCGGGGGTACACAGACA 960

QY 1314 CTAGCGAGTCAGTTCTGTCACACAGTGGCAGGGGTATGACGCTTATCGACGCGCCCTC 1373  
DB 961 CTAGCGAGTCAGTTCTGTCACACAGTGGCAGGGGTATGACGCTTATCGACGCGCCCTC 1020

## RESULT 5

US-09-738-626-2906  
; Sequence 2906, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOL, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 95/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 2906  
; LENGTH: 759  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-2906

Query Match 11.3%; Score 671.8; DB 10; Length 759;  
Best Local Similarity 93.7%; Pred. No. 2.9e-201;  
Matches 711; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1548 ATGGAATCACTATCTGCAAGACGAGCAAGAGTGGCAAGAGTGGCAGTTCGAGTCTTAATC 1607  
DB 1 ATGGAATCACTATCTGCAAGACGAGCAAGAGTGGCAAGAGTGGCAGTTCGAGTCTTAATC 60

QY 1608 GCACCCCTCGCCCAACAAGGTTGGAACCTTGGGCTTTCGAACAGATCCTCACACGTAGT 1667  
DB 61 GCACCCCTCGCCCAACAAGGTTGGAACCTTGGGCTTTCGAACAGATCCTCACACGTAGT 120

QY 1668 ACCTACCAAGAGTTCATTCGATGATGAAGCTGGGGAAGTTCATTCAGAAAGTCAAG 1727  
DB 121 ACCTACCAAGAGTTCATTCGATGATGAAGCTGGGGAAGTTCATTCAGAAAGTCAAG 180

QY 1728 GCATTCCTGTTGGTGAATACGTTGGGACTAACCCGTCGACGATGAAACAGCTACTTTAA 1787  
DB 181 GCATTCCTGTTGGTGAATACGTTGGGACTAACCCGTCGACGATGAAACAGCTACTTTAA 240

QY 1788 ACCATTGCAAGAGTTCATTCGATGATGAAGCTGGGGAAGTTCATTCAGAAAGTCAAG 1847  
DB 241 ACCATTGCAAGAGTTCATTCGATGATGAAGCTGGGGAAGTTCATTCAGAAAGTCAAG 300

QY 1848 GATGGTCAAAACCCCTGATCCATACGAAGAGTTCGAGTATGAGGCAAAAGATCGCTGCA 1907  
DB 1848 GATGGTCAAAACCCCTGATCCATACGAAGAGTTCGAGTATGAGGCAAAAGATCGCTGCA 1907

Db 301 GATGGTCAAAACCTGATCATACGAAGCAGCTGCAGAGTATGAGCAAGATCGCTGCA 360  
QY 1908 GAATCGTTGAAGTTCATTAATCCCTTGGCATCGCGGGAACGG-CACATCCCTTTCATTGAA 1966  
Db 361 GAATCGTTGATGTTTCAATATCCCTTGGCATCGCGGGAACGGCCACATCCCTTTCATGAG 420  
QY 1967 CCATCATCTTCTGTGAGAGTACAAAGGTTCCAGGCGCTGCACCTAAAACCTGTGGAG 2026  
Db 421 CCATCATCTTCTGTGAGAGTACAAAGGTTCCAGGCGCTGCACCTAAAACCTGTGGAG 480  
QY 2027 GACAAGCTGCGATCTTCAACACCATCGAAGAGTCCCAACCCAGCCCTCACCCAGGT 2086  
Db 481 GACAAGCTGCGATCTTCAACACCATCGAAGAGTCCCAACCCAGCCCTCACCCAGGT 540  
QY 2087 TTGGGACATTTGTCCCGCGCGCAAAACATCGTTGGTGGCAACTGGTGAAGAAAAGCC 2146  
Db 541 TTGGGACATTTGTCCCGCGCGCAAAACATCGTTGGTGGCAACTGGTGAAGAAAAGCC 600  
QY 2147 GAGCCATCGCGGGAAGTGTGGAAGCCCGAGTACTGCTTCTTGGCCAGGTTCATCTCG 2206  
Db 601 GAGCCATCGCGGGAAGTGTGGAAGCCCGAGTACTGCTTCTTGGCCAGGTTCATCTCG 660  
QY 2207 TAGATGCACAACATGCACCATCATCGTTGGATGAAGCAGCATCCCAAGCTGGAAGAAC 2266  
Db 661 CAGATGCACAACATGCACCATCATCGTTGGATGAAGCAGCATCCCAAGCTGGAAGAAC 720  
QY 2267 GCTGATCATACCGTCTCATGAGCAATTAAGCTCGGC 2305  
Db 721 GCTGATCATACCGTCTCATGAGCAATTAAGCTCGGC 759

RESULT 6

US-09-746-660A-31/C  
; Sequence 31, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: Patent In Vers. 2.0  
; SEQ ID NO 31  
; LENGTH: 1059  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(1036)  
; OTHER INFORMATION: RXA00044

US-09-746-660A-31  
Query Match 3.6%; Score 215.2; DB 11; Length 1059;  
Best Local Similarity 96.5%; Pred. No. 1.7e-56;  
Matches 220; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 AGTCCGTGAGCGCCACCATTTGATGTGGTGTACCGAGCTTGGGAGGCTTTTACATCT 60  
Db 228 AGTCCGTGAGCGCCACCATTTGATGAGGTGGTCAACGAGCTTGGGAGGCTTTTACATCT 169  
QY 61 AGCTCCCGTGGGCGTGGAGTGGGTCATTACGCGGTGGGATCACCGCGGTGAAAGTTGCG 120  
Db 168 ACATCGCGTGGGCGTGGAGTGGGTCATTACGCGGTGGGATCACCGCGGTGAAAGTTGCG 109  
QY 121 GAACCCATCTGTCTTCTTGTGGTGTAGGGAACGAGTCCGGTGTGAGAGTGTTCAGATG 180  
Db 108 GAAGCCATGGTCTCTTGTGGTGTAGGGAACGAGTCCGGTGTGAGAGTGTTCAGATG 49  
QY 181 TCTGCAGTTTTTAAGTATGATCATCATCAGCTTGGAAAGCTGAGTAAT 228  
Db 48 TCTGCAGTTTTTAAGTATGATCATCATCAGCTTGGAAAGCTGAGTAAT 1

RESULT 7

US-09-974-300-829  
; Sequence 829, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 10085 500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 829  
; LENGTH: 975  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-829

Query Match 3.2%; Score 192; DB 10; Length 975;  
Best Local Similarity 51.4%; Pred. No. 3.5e-49;  
Matches 472; Conservative 0; Mismatches 440; Indels 5; Gaps 1;  
QY 4240 TGTGCTGAGATGATCAACCTGATGCGATCTGCGCGTTCGCGTCTTTCGCGAGTGTGGT 4299  
Db 27 TCTCGCTAATATGATTAACCTTGATCGCGGACTGCTTTCGATTTCTCCCTGCGCTAAT 86  
QY 4300 TGGTTTCCCGCAACCAAGCGTTTCGGTGGCAATGAGTTCCTGGCGCCGCGCATTTGAT 4359  
Db 87 CGGCTGGTGGCGGCTCAAGCGGTTTGGCGGAAATCCGCTGCTCGCATCTGCTCGGCGT 146  
QY 4360 GCGGATGGTGTTCGCCAACCCCTGGTTAAGCGGTACGACGTGGCGCCACCATGACCGCGG 4419  
Db 147 CATGCTGTGCACCCCTGATTTGTTAAACGCGCTGGGATACGCGGCTGCGGAGCAAGCGG 206  
QY 4420 CGAAATGCCAATGTGCTCCCTGTTGGTGTGGATTTGCTCAAGTGTGTACCAGGGCAC 4479  
Db 207 AGAGATTCCTGTCTGGAATTTAATCGGTTTGGAGTGCAGAGGTCGCGCTACCGAGGCA 266  
QY 4480 CGTGTCTCTGTGCTGGTGTCTCTTGGATTCTTGGCAACGATGCGAGAAGTTCCTGCACAA 4539  
Db 267 GGTGCTTCCCATTTGCTCGCTTCTATTGCTGCAAAAATGAGCGCTTTTTCACAAA 326  
QY 4540 GCGACTCATGGGCACTGCAGACTTCTGATCACCCAGTGTGACTCTGCTCTCACCAG 4599

Db 327 GCGACGCCGAGAGCATTACGCTCTGTTGGACACGATTACGCTGCTTGTACAGG 386  
Qy 460 CTTCCTTACGTTCACTGCTATTTGGTCACGCAATGGCTGGGTGGGTGACTTGTGTCACA 459  
Db 387 TTTTGTCTTCAATTTATGCTCTGCTGAGCAGGATTACATTTGCGCAATGTTTAAATC 446  
Qy 466 CGGCTCAGGAGCTATGATTTGCTGCTGAGCAATTCGCGTACTGGGCGGACTGTATACGAGGGCT 4719  
Db 447 AGGCTTATGCTGCTTTGAACAAATTCGCGTACTGGGCGGACTGTATACGAGGGCT 506  
Qy 4720 CTACTACCAATCGTTATCACTGCTGTCACAGTCTCTCCGCGCAATGAGTGGAGCT 4779  
Db 507 TTACGCGCGCTCTGCTATACCGGAATGCATACATCTCTGCGCTGACCTTCAGCT 566  
Qy 4780 GTTC-----AACAGGCTGATCTTCACTTTCGCAACCGCATCCATGCGCAATATCGC 4833  
Db 567 GATCGGCTCAAGCTCGGCGAATTTTATGCGCGATGCTCGCGCTGTCCATATTCG 626  
Qy 4834 GCAGGCTGCAGCATTTTGGCAGTGTCTTCTAGCGAAGAGTGAAGCTCAAGGCGCT 4893  
Db 627 ACAAGGTTACGCGCGCTTGGATGATGTTTATTGTCAAAAGATGAGAAGCAGAAAGCCT 686  
Qy 4894 TGCAGGCTTCAGGCTGCTCCGCTGCTTCTGTTTACAGAGCTCCGATCTTCGCTGT 4953  
Db 687 GTCCCTGACATCCGGAATTTACGCTTATCTCGGAATTTACCGAACCGCATTTGTCGGAGT 746  
Qy 4954 GAACCTTCGCGCTGCGCTGCTTACATTTGATGCTGATCCGACGATATCGGTGGCGC 5013  
Db 747 GAATCTCAGATACAGATTTCCGTTGCTGATCGCGATGATCAGTTCCGGGATTCGCGTAT 806  
Qy 5014 TTTGATGCACTCTTGTATCAAGCAGTGTGCTTGGCGCTGCGAGTTCCTTGGGTGT 5073  
Db 807 GTTCATTTCTTCGCAAGGGTTTGGCAAGCTCTGTCGGCTGCGCGCGCTACCTGGAAT 866  
Qy 5074 GTTCTTATGATGCTCCAGATATGCTATGCTTCTTGTGTTGCGCGTAGTACCTTTGT 5133  
Db 867 TTTCTCTATATGAGCAATCTACTGGGAGCGTTTCCGATCGGAATGCGGATGCTGTTGAT 926  
Qy 5134 CATCGATTCGCGCAGC 5151  
Db 927 GCGCGGTTTCCCGGAAC 944

## RESULT 8

US-08-781-986A-56  
; Sequence 56, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446

; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 56:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30246 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-781-986A-56

Query Match 3.0%; Score 177.6; DB 8; Length 30246;  
Best Local Similarity 49.1%; Pred. No. 1.6e-43;  
Matches 607; Conservative 0; Mismatches 599; Indels 30; Gaps 4;  
QY 3801 AACGCATCTCGCGACATTTGGCGGAGAGCAACATTTGTCGCCGCCGACACTGTGCAA 3860  
Db 10819 AAGCCATCTAACCCTATTGGGGGAAAGAAATCTTGAAGCTGCAACGCATTTGTGTA 10878  
QY 3861 CGCGTTTACGCTCTGCTCAAGACACCAAGGATGTGGATCGCCAAAGTCTTGGATGATG 3920  
Db 10879 CAGGATTACGTTTAGTGTGAAGATGAAGTAAAGTTGATAAGACGCATTAAGTAATA 10938  
QY 3921 ATCCAGATCTGAAGGACACTTTGAACCTGGGGGCAATGTTCCAGATCATCTGCGGCCAG 3980  
Db 10939 ACGCGTTGCTCAAGGGGCACTTTAAAGCAGACCATCAATATCAAAATTTGCATTGTC 10998  
QY 3981 GCGATGTGATCATGTTTCAAGAACTCGATGACGCAACCTCCAAAGACATCGCTGTGT 4040  
Db 10999 GAACAGTCATGAGTGTATAGCAGTTTATGATGAACAGGTGCTCAAGAGCT--T 11055  
QY 4041 CCACAGAGCAGCTCAAGATGTTTGGCTTAACAACCCAACTGGTTCAGCCGTGCTGTGA 4100  
Db 11056 CGAAAGATGAAGCGAAACCAAGCAGCTGCACAAAAGGGAATCCAGTACACCTTTGATCA 11115  
QY 4101 AGGTATTGGCGGACATTTTCGCCCTGATTCCTCAATCTTGGTGGTGGCGCTCTGCTCA 4160  
Db 11116 AATTGTGGGGATATTTTATACCAATATACCTCGATGTTGACAGCTGTTGTTAA 11175  
QY 4161 TGGCTATCAACAATGTTGGTTGCGCAGGATCTGTCGGTCGGCAATCAGTGGTGGA 4220  
Db 11176 TGGGAATCAATAATTTACTTACAAATGAAAGGTTTATTTGCTCCAAAGACATTTAGAGA 11235  
QY 4221 TGTTCCTCAGATCAGCGGTGTTGCTGAGATGATCAACCTGATGATCTGCGCGCTTCG 4280  
Db 11236 TGTATCAACAATGCTGATATTTCAACATCATTAATGATTCGAGTACGGCATTTA 11295  
QY 4281 CGTCTTCCAGTGTGTTGGTTTCCCGCAACCAAGCGTTTTCGGTGGCAATGAGTTCC 4340  
Db 11296 TTTTCTTACCAGCATTAATTTGTTGGAGTAGTATGCGTGTATTGTTGGTAGTCCGATTC 11355  
QY 4341 TGGCGCGCGCATTTGGTATGGCGATGTTTCCCAACCTGGTTAACGCTACGAGCTGG 4400  
Db 11356 TAGGCATAGCTTTAGGTTGATTTTAAATGATCCGCAATAGTATCTCAGTATGTTGG 11415  
QY 4401 CCGCCACCATGACCGGCGCGAAATGCCAATGTGTCCTCTGTTGTTGGATGTTGCTC 4460  
Db 11416 -----CAAAAGGGAATATTCGACGTTGGAATATTTGCTTAGAGATAAGC 11463  
QY 4461 AAGCTGTTTACCAGGCGCGCTCTCTGCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCT 4520  
Db 11464 AGTTGAATTACCAAGGTCAGTGTGCCAGTTTAAATTCAGCTTACGTTTACGTTAGCTA 11523  
QY 4521 TCGAGAAGTTCTCTGCACAAGCGACTCATGGCAGCTGCAGACTTCCTGATCAACCCAGTGT 4580  
Db 11524 TTGAAAAGGATTAAATAAGTCGTTCACGATTTCGATAAAATGTTGCTTGGACCGG 11583  
QY 4581 TGACTCTGCTGCTACCGGCTTCTCTTACGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4640  
Db 11584 TAGCGCTTTTACTGAGTTTATATCATTTGACCATTTTATCATTTGACCATTTGCGTT 11643  
QY 4641 TGGGTGACTTGTCTGGCACACGGTCTGACAGGACTCTATGATTTTCGGTGTGCTCCAGTGG 4700

```
Db 11644 TTGGTACAGGTATATACATCTGGTGTACATTTATATATCCAGCATCGAGTGGCG 11703
Qy 4701 GTCTGCTTTTGGTCTGCTACTACCAATCGTTATCACTGCTGTGACCAAGTCTTCC 4760
Db 11704 GACCATATATGATTGTATATGACCACTTGTATACAGACTACACATATGTTT 11763
Qy 4761 CGCAATATGAGTGGAGCTGTT-----CAACAGGAGTGGATCTTCACTCTCGCAACGG 4814
Db 11764 TAGCAGTAGATTTTCCAAATGATGGGTAGCAGCTTAGCGGTACGATTTATATGCGCAATTG 11823
Qy 4815 CATCCATGGCAATATATCGCGAGGTCAGCATGTTTGGCAGTGTCTTCTCCAGGAAGA 4874
Db 11824 TTGGATTTCCAAATATTTTCAGGCTCTGCAGCATTTGGAGCATGGTTTCTATAAAC 11883
Qy 4875 GT-----GAAAGCTCAAGGCTTTCAGGCTCTTCAGGTGCTCCGCTGTTCTTG 4925
Db 11884 GTCGTAATGGTTAAAGAAAGGCTTGGCATTAACATCTTGTATTTCTGTATGTTAG 11943
Qy 4926 GTATTACAGAGCTCGCATCTTCGCTGTGAACCTTCGCTGCGCTGCGCTTCTACATG 4985
Db 11944 GTGTTACTGAACGAGCATGTTTCGCTGTGAACCTTACCTCTGAAATATCCATTTATCGCTG 12003
Qy 4986 GTATCGGTACCGAGCTATCGTGGCGCTTGATTG 5021
Db 12004 CGATATCAAGCTCTGTGTATTGGGGGCAATCGTTG 12039

RESULT 9
US-09-974-300-685
; Sequence 685, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 685
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-685

Query Match 2.9%; Score 174.2; DB 10; Length 1380;
Best Local Similarity 48.9%; Pred. No. 2e-43;
Matches 524; Conservative 0; Mismatches 608; Indels 45; Gaps 4;

Qy 3779 ATGCACCATAGGACCTCGCGCAACGCATCTCGCGACATTCGCGGAGCAACATT 3838
Db 1 ATGATTCATAAGAAATCGCTGTACGTTTATACAGCTTCTTGGCGGAGAGATATGTC 60
Qy 3839 GTCGCCCGCGCACACTGTCAACCGCTTACGCTGTGCTCAAGACACCAAGGATGTG 3898
Db 61 ATCAGTGCAGCTATTGTGCGCAACAGGCTTCGTTTGGTGATTAAAGGACGCAAAATC 120
Qy 3899 GATCCCAAACTCTGGATGATGATCCAGATCTGAAGCCACCTTTGAACCTGGCGCATG 3958
Db 121 GACGAGGAAAGGTCGAAGAGCTGGAAGGTGTAAGAGGCTTTTCAAGCTCTGGTCAA 180
Qy 3959 TTCAGATCATCTCGCGCCAGCGGATGTGATCATGTTTTCAAAGAACTCGATGAGCA 4018
Db 181 TACCAGATATTTCGGAACAGGCTTGTATTAAGTGT--ATGAACAGTTTGGCAAA 237
Qy 4019 ACCTCCAAAGACATCGCTGTGTCCACAGAGCAGCTCAAGATGTTGTGGCTAACACGCC 4078
```

RESULT 10

```
Db 238 GAGCTGAATCTTGACCAAAAGGAACGGTTCAGCACAGCATGCCGCAAGGGAATATG 297
Qy 4079 AACTGGTTTACCGCTGTGTGAAGTATTTGCGGACATTTTCGTCGCGTGTATTCCAATC 4138
Db 298 AATCCTTTAGCGCATTTTGGAAAACGCTGTCAAATATTTTCGTGCAATCATCCCGCC 357
Qy 4139 TTGTTTGTGGCGGTCTGCTCATGCTATCAACAATGTTGTTGTTGCGCAGGATCTGTC 4198
Db 358 ATTGTTGCAAGCGGTTTATTAATGGGCTGTGGGAATGATGAAGGATTTTCAGTGGGTC 417
Qy 4199 GTCGCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258
Db 418 GATCCTGATGGCGGCTGTA-----TAAATGCTTGTAT 450
Qy 4259 CTGATGGCATCTGCGCGCTTCGCGTCTTCCAGTGTGTTGTTGTTTCAACGCAACCAAG 4318
Db 451 ATGTTTCTCGAGCGCGCATTTATATCTTCCGATCTTGTGCGGTGTCAGTGTGCGGAA 510
Qy 4319 GCTTTTGGTGGCAATGAGTCTCTGCGCGCGCATTTGATGTCGATGTTGTTTCCCAAC 4378
Db 511 GAGTTTGGCAGCAATCTTACTTGGGGCTGTACTTGGAGGAATTTTAAATTCACCGGAGC 570
Qy 4379 CTGGTTAAGGCTACGACGTGSCGCCACCATGACCGGGCGGAAATSCCAATGTGGTCC 4438
Db 571 TTGTTCAATCTTGGGACTTGCARCTGCATCACCAGATG-----TGATGAC 618
Qy 4439 CTGTTTGGTGGATGTTGCTCAAGCTGTTTACCAGGCAACGCTGTTCTGTGCTGTGGTG 4498
Db 619 TTTTACGGACTCAAAATCGAGATGCTTGGCTATCAAGGAACGCTGTTTCCAAATTTGCT 678
Qy 4499 GTCTCTTGGATCTGCGACAGTTCAGAGTTCCTGTCACAGGCTCATGGCACTATGGCACTGCA 4558
Db 679 GCGGTATATGATGAGTAAATCGAAAACGCGTAAAGAAATCGTTTCCAAATCGGATC 738
Qy 4559 GACTTCCCTGATCACCCAGTGTGACTCTGTGCTCAACGCGCTTCTTACGTTTACGTTTAC 4618
Db 739 GATCTATATGTAACACCGTTTGTACGATTTTCAACGGGATTCATTGCTTTTATCGG 798
Qy 4619 ATTGTTCCAGCAATCGCTGGTGGTGTGCTGTCGACAGGCTCTCGAGGACTCTAT 4678
Db 799 ATCGGTCTTTAGGAAGATGCTTGTACAGGATTCACGAACATTTTAAACATATGTGAT 858
Qy 4679 GATTTCCGTGGTCCAGTGGCGGCTGCTGCTTTTCGGTCTGTTCTACTACCAATCTTATC 4738
Db 859 GACGTTCGCGGCTGTTGCGGACTGATTTTCGGGGGACATATCTCGCTGATCTTCTA 918
Qy 4739 ACTGTTCTGCACCACTCTTCCCGCAATGAGCTGAGCTGCTGTC---AACAGGCTGGA 4795
Db 919 ACTGTTCTCCACACAGTTTCCATGCGATTGAGCGCGGCTTATCGCTGATATCGCGAGG 978
Qy 4796 TCCTTCATCTTCGCAACCGCATCTCCATGGCAATATCGCGGAGGTCGACATGTTTGCA 4855
Db 979 AACTACTCTCTCCGATCTGTGCTATGTCCAATGTAGCCAGGCGGGGAGGAAAGCC 1038
Qy 4856 GTGTTCTTCTTACGCAAGAGTGAAGAGCTCAAGGCGCTTTCAGGTCGTTTCAGTGTCTCC 4915
Db 1039 GTATTCTTTTGGCAAGGCTGCAAAACAAAGAAATTCCTTACCTGCCCTTCTCT 1098
Qy 4916 GCTGTTCTTGTATTACAGAGCTTCGATCTTCGCTGTAACCTTCGCTGCGCTGGCGG 4975
Db 1099 TCTTTTCTCGGATTAACCAAGCTGTTATTTTCGGTGTCACTCGTTCGCTACCGCGCGG 1158
Qy 4976 TTCTACATTTGATTCGTTACCGCAGCTATCGGTGGCGCTTTCGATTCGATCTTCATATG 5035
Db 1159 TTTATCGGCAATGSCCGGGGAGCGCTGGAGTGCATACGTTGCTTTTATGATGTG 1218
Qy 5036 AAGGCAGTTGGTGGG 5052
Db 1219 GCAGCAACGCAACGG 1235
```

US-09-815-242-9369

; Sequence 9369, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9369

; LENGTH: 1884

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1884)

; US-09-815-242-9369

Query Match 2.7%; Score 163.6; DB 9; Length 1884;

Best Local Similarity 47.0%; Pred. No. 5.7e-40;

Matches 886; Conservative 0; Mismatches 914; Indels 84; Gaps 9;

QY 3779 ATGGACCAATAGGACCTCGCGCAACGCGATCTGCGCGACATGCGCGGAGAGCAACATT 3838

Db 1 ATGAACAATAGGAAATGCAAAAAAGTCATCGATGCGCTTGGCGGACGTGAATATGTC 60

QY 3839 GTCGCGCGGACACTGTGCAACGCGTTTACGCTCGTCTGCTCAAGACACCAAGGATGTG 3898

Db 61 AATAGTGTGCCACTGTGCGACTGCTACGTGTCTGCTCAAGATGAAGAGAAATC 120

QY 3899 GATCCCAAGTCTGGATGATGATCCAGATCTGAAGGACACTTTGAACCTGGCGGATG 3958

Db 121 AATAAAGAGTGATGAGAACTTGAAAAAGTTCAAGGTGCTCTTTAACTCAGGGCAA 180

QY 3959 TTCCAGATCATGTCGGCGCCAGGCGATGTGATCATGTTTCAAGACTCGATGACGCA 4018

Db 181 TACCAAAATATCTTTGGTACAGGTACAGTTACAAAATGTACGATGAAGTTGTTACTT 240

QY 4019 ACCTCCAAAGACATCGCTGTGTCCACAGAGCAGCTCAAGATGTTGTGGCTTAACAAGCC 4078

Db 241 GCGTTACCAA-----CATCATCTAAGGATGACATGAAGCAGAAAGTTGCTTAACAAGGG 294

QY 4079 ACTGGTTACGCGGTGCTGTGAGAGTATGCGGACATTTGCTCCCGCTGATTCCTCAATC 4138

Db 295 AACTGGTTCCAAAGTCTGATCGTACTTTGCGGTGATGTTTGGTTTCCAAATCATCCAGTT 354

QY 4139 TTGGTTGTGGCGGCTGCTGCTATGCTATCAACAATGTGTTGGTTGCGCAGGATCTGTTTC 4198

Db 355 ATCGTAGGCACAGGCTCTCTTCATGGGTGCGGTGCTTTTCAAGCTCTTGAATGCA 414

QY 4199 GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258  
Db 415 CTTCAG-----GTGACTTTTGAACATTTACACAA 444  
QY 4259 CTGATGGCATCGCGCGCTTCGCTTCTTCCAGATGTTGGTTGGTTTCAACCAACCAAG 4318  
Db 445 ATCTTGACAGATACAGCCTTCATCATCTTGCCAGGTGTTGGTTGGTCAACCTTCGCT 504  
QY 4319 CGTTTCGGTGGCAATGAGTTCTCTGGCGCGCGCATGTTATGCGGATGGTTCCTCAACC 4378  
Db 505 GTATTGGTGGAAATCCTCGCTTGGTATGCTTGGTATGATCTTCTCTCTCTGGCTCA 564  
QY 4379 CTGGTTAAGCGCTACGACGTGGCGCGCAACATGACCGCGGCGGAAATGCCAATGTTGCC 4438  
Db 565 CTTCCAAACGCTTGGCAGTTG-----TCAAGGTGGTGAAGTAAACAGCATGAAC 615  
QY 4439 CTGTTGGTGGTGGATGTTCTCTCAAGCTGTTTACCAGGCGACCGTCTTCTCTGCTGGTG 4498  
Db 616 TTCTTTGGTT---CATCCCTTCTTGGTTTGAAGGTTCGCTTCTTCCAGCCCTCATC 672  
QY 4499 GTCTCTTGGATCTGTGCAACGATCGAGAAGTTCTCTGCAACAGCATCATGGGCACTGCA 4558  
Db 673 ATCGGGTGTCTGGAGCTAAATTTGAAGAAGCTGCGCAAGTTGTTCCAGATGTCATT 732  
QY 4559 GACTTCTGATCACCCAGTGTGACTCTGCTGCTCACCGGCTTCTTACGTTTCACTTGTCT 4618  
Db 733 GACCTCTTGTAAACACCATTCGTGACACTTGTGGTCACTCTATCTTGGACTCTTTGTC 792  
QY 4619 ATGTTCCAGCAATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4678  
Db 793 ATTGGACCAAGTTTCCACGTTGTTGAAAACATACATCTTATTTGCTACAAAAGGATCT 852  
QY 4679 GATTTCGGTGGTCCAGTCCGCGGTCTGCTTTTGGTCTGCTTACTTCAACATCGTTATC 4738  
Db 853 AGCATGCCATTTGCTCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 912  
QY 4739 ACTGGTCTGCAACAGTCTTCCGCGCAATTTAGCTGGA---GCTGTTCAACAGGATGGA 4795  
Db 913 TCAGGTGTGCAACCAATCTTCAACTTGTGTTGAGTGCATTTACTTCTGCTGACCATGCT 972  
QY 4796 TCCTTTCATCTTCGCAACCGCATCCATGGCAATATCGCGAGGTCGACGATGTTTGGCA 4855  
Db 973 AACCCATTCAACGCTATCATCAGCTGCTATGACAGCTCAAGGTCTGCTACTGTTGG 1032  
QY 4856 GTCTTCTTCTAGCGAGAGTGAAGCTCAAGGCTTGCAGGTCTTTCAGGTGCTCTCC 4915  
Db 1033 GTTGGTGTAAACAACAAATCCAACTGAAACACTTTCCTTCCGCGCTCTCTTCT 1092  
QY 4916 GCTGTTCTTGGTATTACAGAGCCTCGATCTTTCGTTGTAACCTTCGCTCGCTGGCGG 4975  
Db 1093 GCCTTCTAGGTATTACAGAGCCTGCTATCTTCGGGTGAACCTTGGCTTCCGTAACCA 1152  
QY 4976 TTCTACATTTGGTATCGGTACCGGATCGGTGGGCTTTGATTTGCACTCTTTGATATC 5035  
Db 1153 TTCTTCTTCAATGATGCTGTGCAATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1212  
QY 5036 AAGCAGTTCGTTGGCGCTGCGAGTTTCTTGGGTGTTGTTTCTATTGATGCTCCAGAT 5095  
Db 1213 GCTGTACTGTAATGATCATCACTCATCCCTGGTGAACATGCTTTATGTTGTAACGGA 1272  
QY 5096 ATGTCATGTTCTTGGTTTGGCGGTAGTTACCTTTTGTCTATCCATTCGCGCAGGATT 5155  
Db 1273 CAAGTCCCAATACCTTCTTATGTTAGTGTATCATTT-----TGCCCTT 1317  
QY 5156 GCTTATGGCTTTACTTGGTTCGCGCAACGGAGCATTTGATCCAGATGCAACCGTGTCT 5215  
Db 1318 GGTTCCTTCTTACTTACTATGTTGGTTTACGAAGA---TGAAGTAGACGCAACTGACGT 1374  
QY 5216 CCAATGCTCGAGGACGCAACAAAGCCGAAGACGACACCCGCGAGATTTTCAACGAT 5275  
Db 1375 GCAAAAACAGCTGAAGTGGTGGTGAAGAAAAAGAAAGTTGCGCAGCAGCTCTTCAAAAT 1434



Db 2049 CACAAGGTGGCGATACGATGCTACCGATGTTATTACCAGCGGTTATTGCACAAGCGGGG 2108  
Qy 4845 CATGTTGGCAGTGTCTTCTAGCGAGAGTGAAGAGCTCAAGGGCTTCGAGGTGCTT 4904  
Db 2109 CTGTTTACGTGCTCTTTTCTCAACAAAATGTGAACATAAAGGTTGGCTTGTCTT 2168  
Qy 4905 CAGTGTCTCCGCTGCTTCTGTTATACAGAGCTCGGATCTCCGGTGTGAACCTTCGCC 4964  
Db 2169 CAAGTATTACGACATATTTTGGAAATCTAGCAACCACTGTATATGGGTGACATTTACCAT 2228  
Qy 4965 TGGCGTGGCGTTCTACATTTGGTATCGGTACCGGAGTATCGGTGGGCTTGTATTCGAC 5024  
Db 2229 TGAATAAACCAATTTATTGACGCTTGTATGCTGGCGTATCGGTGGTGCATTTGGCTA 2298  
Qy 5025 TCTTGTATCAAGGACGCTTCGTTGGGCTGCGAGCTTCTTGGGTGTTCTTCTATG 5084  
Db 2289 TGAATCAGCTGAATAACCTTTACGTTTGGCTTGTAGTATGTTGAGCTTGGCTGGCTTA 2348  
Qy 5085 ATGCTCCAGATATGCTATGTTCTTGGTTTCGCGGGTAGTTACCTTTGTCTATCCGATCG 5144  
Db 2349 TTCTTGCAGACAAAAGATACCTGACCGATGATTACTGTTGGTGAATTTGGTCCGGAATTG 2408  
Qy 5145 GCGGAGGATGCTTATGCGCTTACCTTGTGTTGCTGCCGCAAGCGGAGGATTCATCCAGATG 5204  
Db 2409 CTTTATCATTTGGCTTCTTAACTTGTCTTACGTT-----TTGAAGATCAAC 2459  
Qy 5205 CAACGCTGCTCCAGTGCCTGAGGACGACCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAT 5264  
Db 2460 CTAATCCAGAACAGCACTGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2519  
Qy 5265 TTTCAGAGATTCACCACTATCCAGGCA---CCTTTGACCGGTTGAGCTATGCACTGA 5321  
Db 2520 CGAATCAGAGACAAAATTTATTAGCACTCCACITTCAGGTGAAATTTTACCGTAG 2579  
Qy 5322 CGAGCGTCAGGATGCCATGTTTGGCCAGCGGAGAGTTCGCTCAGGTGTTGCGATCGTCC 5381  
Db 2580 AAAAAGTACAAGACCTGTTTTCAGGTGCTTTTAGGAAAAGGTTTGAATTTGAGC 2639  
Qy 5382 CCACCAAGGCGAGCTGTTTCCAGGTGAGCGGAGAGATCGTGTGGCTTCCATCTG 5441  
Db 2640 CGACTGAGGCAACTGTATGACCCGAGATGTTGAAATCACCACATTTTTCGACAG 2699  
Qy 5442 GTCAGCTTCCAGTCCGCACTAAGGCTGAGGATGTTTCCAAATGTTGATATCTTGATGC 5501  
Db 2700 GACATGCTGTTGGCTTGACGACACAGAGG-----GGCTTGAATTTAATGC 2747  
Qy 5502 ACATTTGTTTCGACACCTTAACCTCAAGCGCAGCAGCTTTAACCCGCTGAAGAGAGG 5561  
Db 2748 ATATTGGATGATAGCTGCAATTTAGATGTTAAGGCTTTGAATTTACGTGAACAG 2807  
Qy 5562 GCGATGAATCAAGCAGGAGCTGCTGTGTGAATTCGATTTGATGCCATTAAGGCTG 5621  
Db 2808 GTGATCTGTTAAAAAGAGAGATTGCTAGTTACTTTTGTATTTGCTGCCATTAAGAG 2867  
Qy 5622 CAGGTTATCAGTATACCGCGGATGTTGTTTGAATTAACA 5664  
Db 2868 CTGTTATCCGCTAGTTACCGGATTTGGTGAACGAATACGAA 2910

RESULT 13  
US-09-070-927A-564  
; Sequence 664, Application US/09070927A  
; Patent No. US20020120116A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles A. Kunsch  
; Steven Barash  
; Patrick J. Dillon  
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 982  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville

STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICANT: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 664:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 567 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 664:  
US-09-070-927A-564

Query Match 2.08; Score 122.2; DB 10; Length 567;  
Best Local Similarity 55.98; Pred. No. 3e-27;  
Matches 271; Conservative 1; Mismatches 209; Indels 4; Gaps 2;  
Qy 3824 GGCGAAGACACATGTCGCGCGGACACATGTCGACGCGGTTTACGCTGCTGCTCAAA 3883  
Db 83 GGAAGATATATCTTATCGCTGTCGACATGTCGACACGCTTGTGCTCGTACTAAAG 142  
Qy 3884 GACACCAAGATGTCGACGCGGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 3943  
Db 143 GATAATCAAAAGTTAATCAAAAGCTTTGGATGAAACCCAAATGTAAGAGGTACCTTT 202  
Qy 3944 GAACTGCGCGCATGTCACAGATCATGTCGCGCGGCGGATGTCGATGATGATGATGATGAT 4003  
Db 203 AAAATTGATGTCAGTATCAGTTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 262  
Qy 4004 GAACTGATGACGCAAC---CTCCAAAGACATCGCTGTGTCACAGAGAGAGAGAGAGAGAT 4060  
Db 263 GAACCTTATAAAAAAAGTTCGCTTTCAGAGCTTTCACAGAGAGAGAGAGAGAGAGAT 322  
Qy 4061 GTTGTGCTACACAGCGCAACTGTTTCAGCGGCTGCTGTGAGGTATTTGGCGGACATTTTC 4120  
Db 323 GATAAAATAAAAAATTCAAATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 382  
Qy 4121 GTCCCGCTGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4180  
Db 383 GTACCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 442  
Qy 4181 GTTCGCGAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4240  
Db 443 ACTGCGCGGCTCTTTTGGACCAAAATCAATCAATCAATCAATCAATCAATCAATCAATCA 501  
Qy 4241 GTTCTGATGATCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4300  
Db 502 ATTTCTCTATGATTCAGCTTATGTCAGCTGACCACTTATGCTGCTGCTGCTGCTGCTGCT 561  
Qy 4301 GGTTT 4305  
|||

```
Db 562 GGTAT 566
RESULT 14
US-10-156-761-3925
; Sequence 3925, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3925
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1143)
US-10-156-761-3925
Query Match 2.0%; Score 121.2; DB 14; Length 1143;
Best Local Similarity 49.0%; Pred. No. 1.1e-26;
Matches 395; Conservative 0; Mismatches 393; Indels 18; Gaps 2;
Qy 515 CCCACGATTGTCCTCCGGTTTATTGATCTTCATATCAGGTGGAAACGGTGGCGCT 574
Db 138 CCACTGGCTGTCCTCCGGTTTTCGATCTGCACACACGCGGGGGCGGCGCTCT 197
Qy 575 TCCTACGGACGACGACGACGAGGAGGATTCACCGGCGATATCACCAGGAAATGGCAC 634
Db 198 CACCTCGGGACGCGGAGGAGGTTTCGCGGGCGTCCACACCCACCGCTGCACGGCAC 257
Qy 635 GACCGTATGTTGCCAAGCATGTTTCGCGCGCGGTGAGCGACTGCGACGAGCGAGTGGGA 694
Db 258 CACCACCGTGTGCTCCACCGTACCGCGGAGATGAGGCTGCTGCGCCAGCGCGGG 317
Qy 695 AAACCTTATTCCTTGTGAAGAGTCTGCTGCGGCGATTCACCTGAGGCGCTTT 754
Db 318 GCTGCTCCGAGCTGGCCGAGGAGGATTCGCGGCGATCCACTTCGAGGGCGCT 377
Qy 755 CATCAACGCTGCGTGTGCTCAAAACCGGATTTTCATTTTCCCGCAACCCAAAC 814
Db 378 CATCTCGCGTGGCGAAGGCGCCCACTCCGAGGAACTGCTGCGGACCCGACCGGC 437
Qy 815 AGATCTTCCGGGTGATCATCGGGGAAAGTTGATCAATTCATACATAGAGGC 874
Db 438 CGAGGTCCGAAGTGTACGCGGCGCGCGGCGGCGGCGGAAAGTGTACGCTGGCCAC 497
Qy 875 GGAACCTGACAATCTTCTGAGTCTCTGCGAGCGAGCGACACCATATTCGTTTC 934
Db 498 CGAAGTCCGGGGCGGATCCCTCCGTCGCTGCGCGAGCAGCGGGTGTGCGCG 557
Qy 935 CTTCGGGACACTGATGACATTTTGATACACTACAGGCGAATTCCTTTGGCTAAGA 994
Db 558 GATCGGCGACAGGACGCGAGTACGAGCAGAGGTTGGAGGCGCATCGACCGGGT----- 612
Qy 995 GAAAAATGTGACGCTACGCGATTTGTTCAATGCGATGCTCTCCGCTGCATCATAG 1054
Db 613 -----CGGACGGTGGCCAGCACTGTTCACGCGATGCCCGCGCTCGGCCACCG 662
```

```
Qy 1055 GGCTCCCGGACGCGTGGCGCTTSCHTGTCTGCGGACAGTSCCGGGGACGACATATGTGA 1114
Db 663 CACGCGGGCGGATCGCGCGCTCTGGAGGACGAGCGGATCACGGTTCGAGCTGATCAA 722
Qy 1115 GTTGATCCCGACGCGGTGCAATTTGGC---CGATGGACGCTGATGCTGCTTCCAA 1171
Db 723 CGAGGTACGCATCTGCACCGCGCGCTCCAGCTGGCGTTCCATCACGGGCGCGGG 782
Qy 1172 CAACGCTTTTTCATCACGACGACGATGAAAGCGCGCGGAAATGCCAGACGGTGTAGAT 1231
Db 783 CCGGTGCGCTTTCATCACGACGACGATGAGCGCGCGGGCTTCGGCGACGCGCGCTATCT 842
Qy 1232 TTTGGCGTTTGAACGCTCACCGTCAACGATGGAGTGCCTGCTGCGGATGCGGCGG 1291
Db 843 GCTCGGCGCGTGGAGTGCAGTCAAGGCGGTGCGCGGCTGATGGAGGCGGCTTC 902
Qy 1292 CATCGCGGGGCGACACGACACTAG 1317
Db 903 GATCGCGGCTCGAGGCTCACGCTGG 928
RESULT 15
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 2.0%; Score 121.2; DB 14; Length 9025608;
Best Local Similarity 49.0%; Pred. No. 8.4e-24;
Matches 395; Conservative 0; Mismatches 393; Indels 18; Gaps 2;
Qy 515 CCCACGATTGTCCTCCGGTTTATTGATCTTCATATCAGGTGGAAACGGTGGCGCT 574
Db 4865055 CCACGCGTGTGTCCTCCGGCTTCGTCATCTGCACCAACACGCGCGCGCGCTT 4865114
Qy 575 TCCTACGGGACGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 634
Db 4865115 CACCTCGGCGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4865174
Qy 635 GACCGTATGTTGCCAAGCATGTTTCGCGCGCGCTTCAGCGACTGCGAGCGCAGTGA 694
Db 4865175 CACCAACCGTGTGCGCTCCACCGTCCACCGTCCACCGTCCACCGTCCACCGTCC 4865234
Qy 695 AAACCTTATTCCTTGTGTAAGAGTCTGCTGTCGCGCATTCACCTCGAGGGCGCTTT 754
Db 4865235 GCTGCTCTCCGAGTGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4865294
Qy 755 CATCAACGATGCGCTGTGTTGCTCTCAAAACCGGATTTTCATTTTCCCGCAACCCAA 814
```

```
Db 4865295 CATCTCCCGTGGCGGAAGGCGCCACCTCCGAGGAACCTGCTCCGCGGACCCGCGCCGGC 4865354
Qy 815 AGATCTTGGCCGGGTGATCATCGGGGAAAAGTTGGATCAAAATCATCATACAGTAGCGCC 874
Db 4865355 CGAGGTCCGCAAGCTGATCAGCGCGCGCGGCGAGGAGATGTCACGCTGGCCAC 4865414
Qy 875 GGAACCTGACAATCTTCTCAGCTTCTCGATCTCTCGGAGCGCACCATCATATTGCTTC 934
Db 4865415 CGAATCCCGGGGGCATCCAGCTCCGTCACCCCTGGTGGCGGACACGGGTGATCGCCG 4865474
Qy 935 CTTGGGCACACATGATCAGATTTTGATACACTACACGCGCAATTGCCTTTGGCTAAAGA 994
Db 4865475 GATCGGCACACGACCGACGCTACGAGCAGCGGTGGAGGCCATCGACCGGGT----- 4865529
Qy 995 GAAATGTGACGCTACGCTACGCTAGCGATTGTTCAATGCGATGCCCTCCGCTGCATCATAG 1054
Db 4865530 -----GCGACGGTCGCCACGACCTGTTCAACGCGATGCCCGCGCTCGGCCACCG 4865579
Qy 1055 GGCTCCCGGACGCTGGCGCTTTGCTTGTGGGCGACGTGCCGGGACGCATATGTTGA 1114
Db 4865580 CACGCCGGCGCGATCGCCGCCCTCTGGAGGACGAGCGGATCACGGTCGAGCTGATCAA 4865639
Qy 1115 GTTGATCGCCGACGCGCTGCAATTGGC---CGATGGAACGGTCGATCTAGCTGTTCCAA 1171
Db 4865640 CGACGGTACGCTGTCACCGCGCGCCCTCCAGCTGGCGTTCCATCACCGGGCGCCGG 4865699
Qy 1172 CAACGCCCTTTTTCATCAGGACGCCATGGAAGCGCGCGGAATGCCAGACGGTGAGTACAT 1231
Db 4865700 CCGGTCGCGTTTCATCAGGACGCGGATGGACGCGCGGGGCTTCGGCGACGCGCGCTATCT 4865759
Qy 1232 TTTGGGGGTTTGAACGTCACCGTCCCGATGGAGTGGCGCGCTGCGCGGATGGCGCGC 1291
Db 4865760 GCTCGGCCCGCTGGAGGTCGAGGTCACTGAGGGCGTGGCGCGGCTGATGAGGGCGGTT 4865819
Qy 1292 CATCGCGGGGGCACACACTAG 1317
Db 4865820 GATCGCGGCTCGACGCTCAGCTGG 4865845
```

Search completed: October 1, 2003, 05:00:44  
Job time : 968 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 04:43:20 ; Search time 34 Seconds  
(without alignments)  
1869.632 Million cell updates/sec

Title: US-10-019-284a-2  
Perfect score: 3342  
Sequence: 1 MDHKLQAQRILRDIGGEDNI.....IEAGANLNLNVAKEAVPATP 661  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1609.5	48.2	651	2 S44257	phosphotransferase
2	1478.5	44.2	664	1 B32243	phosphotransferase
3	1227	36.7	480	2 S39978	scrA protein - Sta
4	1178	35.2	480	2 D90038	PTS system, sucros
5	1038.5	32.9	627	2 H96951	fusion, PTS system
6	1093	32.7	479	2 F82432	PTS system, sucros
7	1069	32.0	479	2 J00781	sucrose uptake pro
8	1013.5	30.3	470	2 H83926	PTS system, trehal
9	999	29.9	470	2 C69725	phosphotransferase
10	990.5	29.6	632	2 S68599	phosphotransferase
11	960	28.7	627	2 F95200	PTS system IIAc c
12	957	28.6	617	2 AB1167	phosphotransferase
13	956	28.6	627	2 E98067	phosphotransferase
14	949	28.4	635	2 C95220	trehalose PTS syst
15	934	27.9	705	2 A99084	phosphotransferase
16	933.5	27.9	475	2 C89813	phosphotransferase
17	930.5	27.8	630	2 H83686	hypothetical prote
18	929.5	27.8	636	2 C83724	PTS system, beta-g
19	927	27.7	633	2 AC1436	PTS system, beta-g
20	924.5	27.7	634	2 AD1078	PTS system, beta-g
21	901	27.0	609	2 I40406	beta-glucoside per
22	896	26.8	631	2 B42603	beta-glucoside-spe
23	887	26.5	609	2 T47097	hypothetical prote
24	860.5	25.7	628	2 D97073	PTS system, beta-g
25	859	25.7	617	2 AC1421	beta-glucoside-spe
26	829	24.8	618	2 AC1204	phosphotransferase
27	824	24.7	625	2 C23577	phosphotransferase
28	822.5	24.6	636	2 D86807	hypothetical prote
29	816	24.4	456	2 S62331	phosphotransferase

ALIGNMENTS

RESULT 1

S44257  
phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - Pediococcus per  
C:Species: Pediococcus pentosaceus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 16-Jul-1999  
C:Accession: S44257  
Riekenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.  
submitted to the EMBL Data Library, April 1994  
A:Description: The sucrose and raffinose operons of Pediococcus pentosaceus PPel.0.  
A:Reference number: S44252  
A:Accession: S44257  
A:Molecule type: DNA  
A:Residues: 1-651 <LEE>  
A:Cross-references: EMBL:Z32771; NID:G493728; PIDN:CAA83668.1; PID:G475968  
C:Genetics:  
A:Gene: scrA  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase  
C:Keywords: phosphotransferase  
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III hc

Query Match	48.2%	Score	1609.5	DB 2	Length	651												
Best Local Similarity	49.6%	Pred. No.	1.8e+101															
Matches	331	Conservative	114	Mismatches	196	Gaps	8											
Qy	1	MDHKDLAQRILRDIGGEDNI	VAAAHCA	TRLRRLVLDK	TRDVRQ	SLDDDD	PD	LKGF	TGGM	60								
Db	1	MNHQVADRVLNAI	-GKNNIQA	AHCA	TRLRRLVLDK	TRDVRQ	SLDDDD	PD	LKGF	59								
Qy	61	FOIIVGPGD	VDVFKELDD	ATSKDIA	VSTEQ	LKD	YVA--	NNANW	FSRA	KVLADIF	PLI	118						
Db	60	YQIIVGPGD	VDVFKELDD	ATSKDIA	VSTEQ	LKD	YVA--	NNANW	FSRA	KVLADIF	PLI	118						
Qy	119	PLIVGGGLM	LNINVLVA	QDLFG	PSQSL	VMFP	QISG	VAEM	INLM	ASAP	FAEL	PLV	LGFTA	178				
Db	119	PALVAGGLM	LNINVLVA	QDLFG	PSQSL	VMFP	QISG	VAEM	INLM	ASAP	FAEL	PLV	LGFTA	178				
Qy	179	TRFSGN	FLGAG	ICNANW	FTLV	NGYD	VAAT	WTAG	EMPM	SLF	GDL	VAQ	YOG	TVLPV	238			
Db	179	TRFSGN	FLGAG	ICNANW	FTLV	NGYD	VAAT	WTAG	EMPM	SLF	GDL	VAQ	YOG	TVLPV	238			
Qy	239	LVVSVILAT	IEFLHK	RLMG	TADFL	ITP	VLIT	LTG	FLTA	IG	IPAM	RW	GDL	LH	AGJQG	298		
Db	239	LVVSVILAT	IEFLHK	RLMG	TADFL	ITP	VLIT	LTG	FLTA	IG	IPAM	RW	GDL	LH	AGJQG	298		
Qy	299	LYDFG	PGVGL	FLGV	LVSP	IVIT	GLHQ	SFP	PIEL	ELF----	NOGS	FT	ATAS	MANIA	QOG	354		
Db	299	LYDFG	PGVGL	FLGV	LVSP	IVIT	GLHQ	SFP	PIEL	ELF----	NOGS	FT	ATAS	MANIA	QOG	354		
Qy	355	AACLA	VEFLAK	SEK	LEK	LAG	ASVS	AVL	GITE	IPAT	FGV	NLR	LRW	PEY	IGIT	AAIGG	ALI	414
Db	359	AATL	FAIT	ATK	SQK	OKALT	SSAG	SALL	GITE	IPAT	FGV	NLR	LRW	PEY	IGIT	AAIGG	ALI	418



RESULT 4  
D90038  
PTS system, sucrose-specific IIBC component [imported] - Staphylococcus aureus (strain N  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C/Accession: D90038  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oqu  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsum, K.  
Lancet 357, 1225-1240, 2001  
A>Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; MUID:21311952; PMID:11418146  
A/Accession: D90038  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-480 <KUR>  
A/Cross-references: GB:BA000018; PID:g13702328; PIDN:BA843469.1; GSPDB:GN00149  
A/Experimental source: strain N315  
C/Genetics:  
A/Gene: scrA  
C/Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 35.2%; Score 1178; DB 2; Length 480;  
Best Local Similarity 50.4%; Pred. No. 2.3e-72;  
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;  
QY 1 MDHKLDAQRILRDIGGEDNIVAAAHCAATRLRLVLTQDKVDROSLLDDDDDLKGTFTETGM 60  
DB 1 MNYKQSAEDILNAGIGGENLDMAHCAATRLRLVLTQDKVDROSLLDDDDDLKGTFTETGM 60  
QY 61 FOIIVGPDVHVFKELDDATSKDIAVSTEQKLD---VVANNANWFSRAVKVLADIFVPLPI 120  
DB 61 YQIIIGSGTVNKVSELEKLTGKE-ASTTSEVRAQSAKNMNPLOREVKMLSDIFVPLPIA 119  
QY 121 LVGGGLMALNNVLAQDL-FGQSLVEMFPQISGVAEMINLMASAPPAPFLPVLTGFTAT 179  
DB 120 IVAGGLMGLNLTAKDLFFSKSLIDVYSQFAGLAEMINVPANAPFTLLPLILIGFSA 179  
QY 180 KRFEGNEFLGAGIGMAVFTLVNGYDVAATMTAGE-MPMWSLFGDLVDAQAGYQGTVP 238  
DB 180 KRFEGNFFLGNLGMILVHLSLMSADDFPKAVEAGRAIPYDWFGLHINQVGYQGVLP 239  
QY 239 LVVSWIILATIEKFLHKLMTADPLTPVLTLLTGLTFTIAGPAMRWYGDLLAHGLQ 298  
DB 240 LVAAIYILASIEKGLRKVPTVLDNLTPLLSIFITAPLTFESFVGPITRQLGYLSDGLT 299  
QY 299 LYDFGPGVGLLGLVYSPVITGLHOSFPPIELF---NOGGSFIFATASMANIAQ 354  
DB 300 LYFEGGALGGLIFGLLAPVITGMHSHFIAVETTLIADATKGGSPFIPIATMSNVAQ 359  
QY 355 AACLAVFPLAK-SEKLGKLAGASGVAVLQITPAIFGVNLRWLPYIGTAAIGGAL 413  
DB 360 GAAIAAFIILKQNKLAGVSAAGISALLGITFAMFGVNLKRYPPFIGAIVSGIGSAY 419  
QY 414 IALFDIKAVLGAAGLGVWSIDA--PDWVNLCAVVTVFIAFGAAIAYGLVLYRNGS 471  
DB 420 IAFKVKAIALGTAGLPGFISINPVHAGWLHYFVGMTISFII----AITVLLLSRXKAN 475

RESULT 5  
H96951  
fusion, PTS system, beta-glucosides specific IIBC component [imported] - Clostridium ad  
C/Species: Clostridium acetobutylicum  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C/Accession: H96951  
R;Nolling, J.; Brennon, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4836, 2001  
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: H96951  
A/Status: preliminary

A/Molecule type: DNA  
A/Residues: 1-627 <KUR>  
A/Cross-references: GB:AB001437; PIDN:AAK78403.1; PID:g15023277; GSPDB:GN00168  
A/Experimental source: Clostridium acetobutylicum ATCC824  
C/Genetics:  
A/Gene: CAC0423  
C/Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfere  
Query Match 32.9%; Score 1098.5; DB 2; Length 627;  
Best Local Similarity 36.8%; Pred. No. 7.7e-67;  
Matches 236; Conservative 122; Mismatches 243; Indels 41; Gaps 8;  
QY 1 MDHKLDAQRILRDIGGEDNIVAAAHCAATRLRLVLTQDKVDROSLLDDDDDLKGTFTETGM 60  
DB 1 MDINKTAKDILKLGGEKNVSAHAATRLRLVLTQDKVDROSLLDDDDDLKGTFTETGM 60  
QY 61 FOIIVGPDVHVFKELDDATSKDIAVSTEQKLD---VVANNANWFSRAVKVLADIFVPL 117  
DB 61 YQIIIGSGTVNKVYKAFVEGT---GISSELSDTKKAAMKNMNLFEFARMLSNIFVPI 116  
QY 118 IPIVGGGLMALNNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPPAPFLPVLTGFT 177  
DB 117 IPIAVASGLMLGLGLDADFHLVNSK-----SGLYVILNMFNSNAFQFLPMIAFS 167  
QY 178 ATKRFEGNEFLGAGIGMAVFTLVNGYDVAATMTAGE-MPMWSLFGDLVDAQAGYQGTVP 237  
DB 168 AAREEKTNPYLAALGAIMIHDPDLQNAW---TLGEGIKHTINIFGLNIGMVGYQGTVP 223  
QY 238 LVVSWIILATIEKFLHKLMTADPLTPVLTLLTGLTFTIAGPAMRWYGDLLAHGLQ 297  
DB 224 ILISVWVMSYIEKGLRKIVPEALDILLTPTLMTITFFAMVGVIGGGRFVDELSLGLQ 283  
QY 298 GLYDFGPGVGLLGLVYSPVITGLHOSFPPIELFELFNQGG---SFIFATASMANIAQ 354  
DB 284 TLYNTGTFSGVLEFGGLYSLIVITGIHSHFAIEAGLLANPAIHKNFLLPIWSMANVAQ 343  
QY 355 AACLAVFPLAKSEKLGKLAGASGVAVLQITPAIFGVNLRWLPYIGTAAIGGALI 414  
DB 344 GAALAVYFETRKMKSTAAAPSFCLLGITPAIFGVNLRVTKFPIAGALGAGIGGYI 403  
QY 415 ALFDIKAVLGAAGLGVWSIDA-PDWVNLCAVVTVFIAFGAAIAYGLVLYRNGSIDP 474  
DB 404 VFTKVAMTAVGTGPIGIAIVKGGFLNY---IIAMILAFGGAFIIMVLGKEITEE 459  
QY 475 DATAVPVAGTTKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASGLSGVAIV 534  
DB 460 DLNKETVNDKIKVEEVES-----VSPVNGVYLLKNVPDKTFAEGLIGDVGVD 509  
QY 535 PTKGQLVSPVSGKIVVAFPPSHAFVTRKAEDGSNVVDILMHIGFDTVNLGTHFNPLKKQ 594  
DB 510 PEDGEVVSIDGTVVHVVFETKHAIAKSK---NGVEMLIHIGIDTVKMEGNGEKSFIN 565  
QY 595 GDEVKAGELLCEDDAIKAAGYEVVTPPIVWSNYKKTGPVNT 636  
DB 566 GEEVKGKDKLIQDLDLVKEKAVSPVILVTIVTNHEDMGFVNS 607

RESULT 6  
F82432  
PTS system, sucrose-specific IIBC component VCA0653 [imported] - Vibrio cholerae (str  
C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C/Accession: F82432  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qiu, H.; Dragoi, I.; Sellers  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A>Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: F82432  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-479 <HEI>

A:Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype EI Tor

C:Genetics:

A:Gene: VCA0653

A:Map position: 2

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 32.7%; Score 1093; DB 2; Length 479;

Best Local Similarity 47.8%; Pred. No. 1.3e-66;

Matches 222; Conservative 85; Mismatches 151; Indels 6; Gaps 3;

QY 1 MDHKLQRIILRDIGGEDNIVAAAHCAATRLRLVLKDTKVDYDQSLDDDDPDLKGTFTTGM 60

DB 1 MDYPIAKQLLESIGKSNIOALAHCAATRLRLVLNDETQINESALQGVKGQFKVAG 60

QY 61 FOIIVPGDVHVFKELDATSKDIATVTEOLKDVVANNANWFSRAVKVLADIIVPLPI 120

DB 61 YOLIFSGIVNQVYAEMAKLFGI-VEMSTNDVASAGAEKQNAQRAVGLSDIFVRIIPA 119

QY 121 LVGGGLMAINNVVLAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTAT 179

DB 120 IVAGGLMLGNLTVAPGLFIEGQSLIDANPCLADLASMINTFANAPFVLPVLLAFSAS 179

QY 180 KFGNEFLGAGIGAMVFTLVNGYDYAATWTAGEMPWMSLFGLDVAQAGYOGTVLPVL 239

DB 180 RXFGNPFUGAALGMLVHPDNLNGWFGSGSVGTPTNVLGFEIEKVGQGSVLPLV 239

QY 240 VVSWILATIEKFLHRLMGTADELITPTVLTLLTGLTFIAIGPAMRWVGLLAHGLQL 299

DB 240 VSAYILAKIENGLRKIVPSVDNLLTPTMLAIFITGLFTFTVVGLTRDVGFMGLDALNWL 299

QY 300 YDFGPGVGLLGLVYSPVITGLHQSPPIELF----NOGGSFIATASMANIAOGA 355

DB 300 YDSAGFVGALFGFIYAPVITGMHSHFATETQLLADIVTGGFTFPIPAAMSNIAOGA 359

QY 356 ACLAVFFLAKSEKLAGAGSVAVLGIETEPAFIENVLRWPFYIGIGTAAIGGALIA 415

DB 360 AALAVGVMTKELKGVAPSGVTALLGITEPAMEGVNKLKRYPIAICGAALASAFIT 419

QY 416 LFDIKAVAGAGFLGVVSIDAPDMVFLVCAVTVFAFGAAI 459

DB 420 LFNVAQALGAAGLPGIISINPOQIGYINGMAISFVAFAALT 463

RESULT 7

QJ0781

sucrose uptake protein - Vibrio alginolyticus

N:Alternate names: enzyme II-sucrose protein

C:Species: Vibrio alginolyticus

C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 18-Jun-1999

C:Accession: JQ0781

R:Blatch, G.L.; Scholle, R.R.; Woods, D.R.

Gene 95, 17-23, 1990

A:Title: Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose uptake-end

A:Reference number: JQ0781; MUID:91071601; PMID:2174811

A:Accession: JQ0781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <BLA>

A:Cross-references: GB:M76768; GB:M30194; NID:g155261; PIDN:AAA27555.1; PID:g155262

C:Genetics:

A:Gene: scrA

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 32.0%; Score 1069; DB 2; Length 479;

Best Local Similarity 46.1%; Pred. No. 5.6e-65;

Matches 216; Conservative 94; Mismatches 151; Indels 8; Gaps 5;

QY 1 MDHKLQRIILRDIGGEDNIVAAAHCAATRLRLVLKDTKVDYDQSLDDDDPDLKGTFTTGM 60

DB 1 MNPYPAKELLTLGKSNITLALHCAATRLRLVADEQKIDEGADINLECVKGQFKVAG 60

QY 61 FOIIVPGDVHVFKELDATSKDIATVTEOLKDVVANNANWFSRAVKVLADIIVPLPI 120

Db 297 FTSIFAAPLVGGLYGLVAPLVVGMHTFLAVDLQLIGTGTFGLPILVLSIAQ 356  
 Qy 355 AACLAVFFLAKSKLGLAGASGVSAVLGITEPAIGVNLRLRWPVYIGTAAIGGALI 414  
 Db 357 SAALAMWFAVRDEKLGSLUSSAVSAYLGLTEPAMFGVNRKFPFVCAIIISAAIGGAFI 416  
 Qy 415 ALFDIKAVAGAGFLGVWSIDAPDMVMFLVCAVAVTFVI 453  
 Db 417 TVNGVLANSIGVGLGPIRFSIQAGWGVFFIGMWIAFIL 455

## RESULT 9

C69725  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: C69725; S67929; J05037; I40497; S67864  
 R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton  
 C.; Broc, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gall  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 R:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Segliguchi, J.; Sekowska, A.; Seron  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: C69725  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-470 <KUN>  
 A:Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CABL12609.1; PID:g2633104  
 A:Experimental source: strain 168  
 R:Helfert, C.; Gotsche, S.; Dahl, M.K.  
 Mol. Microbiol. 16, 111-120, 1995  
 A:Title: Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by a phospho  
 A:Reference number: I40497; MUID:95379486; PMID:7651129  
 A:Accession: S67929  
 A:Molecule type: DNA  
 A:Residues: 324-362, 'L', 364-464, 'G', 466-470 <HEL>  
 A:Cross-references: EMBL:X80203; NID:9580941  
 R:Schoeck, F.; Dahl, M.K.  
 Gene 175, 59-63, 1996  
 A:Title: Analysis of DNA flanking the treA gene of Bacillus subtilis reveals genes encod  
 A:Reference number: J05037; MUID:97074649; PMID:8917076  
 A:Accession: J05037  
 A:Molecule type: DNA  
 A:Residues: 1-139, 'S', 'L', 141-362, 'L', 364-464, 'G', 466-470 <SC2>  
 A:Cross-references: EMBL:Z54245; NID:gl000450; PIDN:CA91014.1; PID:gl000451  
 C:Comment: This enzyme functions as the specific trehalose transporter. It belongs to th  
 C:Genetics:  
 A:Gene: treP, treB  
 C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II  
 C:Keywords: phosphotransferase; sugar transport system  
 F:113-131/Domain: transmembrane #status predicted <TM1>  
 F:160-181/Domain: transmembrane #status predicted <TM2>  
 F:184-203/Domain: transmembrane #status predicted <TM3>  
 F:230-245/Domain: transmembrane #status predicted <TM4>  
 F:263-286/Domain: transmembrane #status predicted <TM5>  
 F:305-325/Domain: transmembrane #status predicted <TM6>  
 F:375-395/Domain: transmembrane #status predicted <TM7>  
 F:402-421/Domain: transmembrane #status predicted <TM8>

Query Match 29.9%; Score 999; DB 2; Length 470;  
 Best Local Similarity 43.9%; Pred. No. 3.1e-60;  
 Matches 203; Conservative 93; Mismatches 160; Indels 6; Gaps 4;

Qy 7 AQRILRDIGEDNVAIAAHCAATRLRLVLKDTKDVDROSLDDDDPDLKGTFTTGGFQIIVG 66  
 Db 8 ARQIVEAGGAAGAAATACVTRLRFRALIDESKVDMLDQIDVYVKGFSFNGOFQVYIG 67  
 Qy 67 PGDYDHPKELDDATSKDIAYST-EQLKDVVANNANWFSRAVKVLADIFVPLIPILYGGG 125  
 Db 68 QGTVNKYIAELVKETG--IGESTKDEVKASEKNPNPLQRAVKRLADIFILPAIVTAG 125  
 Qy 126 LLMAINNVLVAQDL-FGQSLVEMPPQISGVAEMINMASAPPFLPVLVGTFTTKRFGG 184  
 Db 126 LLMGINNLTABGIEFTKTSIVQVYPQWADLANMINLTAGTFTLPALIGVSAVKRFGG 185  
 Qy 185 NEFLGAGIGMAMVFTLVNGYDVAATMTAGEMWMSLFGDLVAOAGYOGTVLPVLVWSVI 244  
 Db 186 NELLGVLGVGLVHVDLLNMGYGAEOSEGIPIVWNLFGLEVQKVGQYGVQLPILLASYM 245  
 Qy 245 LATIEKFLKRLMGTAFLITPVLTLTLLTGFUTFAIGPAMRWGVDLLAHGLOGLYDPGG 304  
 Db 246 LAKIEVFLTKRTPGEGQLLVVAPITLLTGFASFIIIGPITFAIGNVLTSLISVFGSFA 305  
 Qy 305 PVGGLLGLVYSPVITGLHQSFPPTELELQ--GGSFIFATASMANIAOGAACLAVFF 362  
 Db 306 ALGGLLYGFGYSAVLTGMHHTFLAVDQLIGSKLGGFLPMLALSNIAGSALAMMF 365  
 Qy 363 LAKSEKLGLAGASGVSAVLGITEPAIGVNLRLRWPVYIGTAAIGGALIALDIKAV 422  
 Db 366 IVKDEKQKGLSLTSGISAVLGITEPAIGVNLRYRFPFIAMVSSGLAGMYISSQGVLAS 425  
 Qy 423 ALGAAGFLGVSDADPMVFLVCAVTVFVAFGAIAVGLY 464  
 Db 426 SVGVGVPGIFSIMSOYWGAFAGMAIVLIVFPAGTYIARF 467

## RESULT 10

S68599  
 C:Species: Streptococcus sobrinus  
 A:Variety: strain 6715  
 C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 07-May-1999  
 C:Accession: S68599  
 R:Chen, Y.Y.M.; Lee, L.N.; LeBlanc, D.J.  
 Infect. Immun. 61, 2602-2610, 1993  
 A:Title: Sequence analysis of scrA and scrB from Streptococcus sobrinus 6715.  
 A:Reference number: S68598; MUID:93273516; PMID:8500898  
 A:Accession: S68599  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-632 <CHE>  
 A:Cross-references: EMBL:L06791  
 C:Genetics:  
 A:Gene: scrA  
 C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfer  
 C:Keywords: phosphotransferase; sugar transport system  
 F:480-632/Domain: phosphotransferase system glucose-specific enzyme II, factor III h

Query Match 29.6%; Score 990.5; DB 2; Length 632;  
 Best Local Similarity 35.5%; Pred. No. 1.7e-59;  
 Matches 236; Conservative 114; Mismatches 270; Indels 45; Gaps 12;

Qy 1 MDHKDLAQRILRDIGEDNVAIAAHCAATRLRLVLKDTKDVDROSLDDDDPDLKGTFTTGGF 60  
 Db 1 MDNKQIAKEVTEALGGRDNRVSRVAHCATRLRVVVDEAKIDKRAENIDKVGAFNSGQ 60  
 Qy 61 FQIIVGPDVDHVPKELDD---ATSKDIAYSTQLKDVVANNANWFSRAVKVLADIFVPL 117  
 Db 61 YQIIFGTGTNKKIYDEVYDLGLPTS-----STGEQKQEAAGAAQGNWFQMRSTGDFVPI 115  
 Qy 118 IPILVGGGLLMAINNVLVAQDLFGPQSLVEMFPQISGVAEMIN-----LMASAPFAF 169  
 Db 116 IPIVLATGLFNLRLGLTNTDTFLG-----PFGASSKIDINAFNLTLYQVLTDTAF 166



Db 61 LOVIIGNVGVYKALGSFTKLTDDGSEIAKGT---KD-----SDGNFLSKAIDVISGIFT 114  
QY 116 PLIPIVLGGGLMANNVLAODLFGPOSIVEMFQISGVAEMINLMASAPFAFLPVLVG 175  
Db 115 PILGALAGGMLKGLMLITT---FG-----WLTSSGTQIILYAAADSVEYFLPLILA 165  
QY 176 FTRATRGNEFLGAGIGMAMVFPVLVNGYDVAATAGEMPMSLFLGLDVAQAGYQGTV 235  
Db 166 YTAARKFGANPPVATAAGALVYPTMINLFNEGAHITFLOIP-----VVLMSYSFSV 217  
QY 236 LPVLVYSMLATIEKFLHKLMTADFLITPVLTLTLGLTFLFIAIGPAMRWGDLIAHG 295  
Db 218 IPILAVWFLSILERFLNSKIHEAAKTEFLTPMCLMLIVPLTFLAFGLGTFLISQCLASG 277  
QY 296 LOGLDYFGPVGGLLGLVYSPIVITGLHOSPPTEL-ELFNQGSFIFATASMANIAQG 354  
Db 278 YFTINLSPIVAGAPWGAQVQVVLVIFGHWGVPVPTMINLSRYGDTMIAMVPSNFAQA 337  
QY 355 AACLAVFFIAKSEKLGLAGAGSVAVLGITEPAIFGNLRLRWPFFYIGIGTAAIGGALI 414  
Db 338 GASLGVFLTKKPEVKAIAGSAALGFFGITEPSIYGVTLKYKKPFVIASTAGAGIAV 397  
QY 415 ALFDIKAVAGALGAAGFLGVVYSIDAPDV---MFLYCAVVTFTVAFGAATAY-----GLYIV 466  
Db 398 -----GAAGSSGAANA-IPGLILPFIIGKGVFGFIL--GIAYAILSAIGTYFF 444  
QY 467 RRGSIDDPATAAPVPAGTAKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDFAMFAGK 526  
Db 445 ---GYKENADGI---APTKEKETGVE---AEVIVSPRGVIVPLNEVKDEAFSAGL 494  
QY 527 LSGVAIVTKGOLVSPVSGKIIVAPPSGHAFAVRTKAEDGNSVDILMHIGDVTYNLNGT 586  
Db 495 LKGVAIVPQEGKLISPVNGTITETAPTGHAGIRS-----DKGVELLHVGFDTVOLNGK 550  
QY 587 HENPLKKQDEVKAGELLCEFDIDAIAKAGYEVTTPIVYSN 627  
Db 551 YFKLLVAQGRVLVVGQALFEFLDEIAKADGYDITTPIVVTN 591

RESULT 13  
E98067  
phosphotransferase system enzyme II (EC 2.7.1.69) scra [imported] - Streptococcus pneumo  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C:Accession: E98067  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: E98067  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAL00370.1; PID:g15459232; GSPDB:GN00174  
C:Genetics:  
A:Gene: scra  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase  
C:Keywords: phosphotransferase

Query Match 28.6%; Score 956; DB 2; Length 627;  
Best Local Similarity 33.3%; Pred. No. 3.6e-57;  
Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;  
QY 1 MDHXDLAQRILRDGGEDNIVAAACATRLRLVLKDKVDVDRQSLDDPDLKGTFFETGGM 60  
Db 1 MNQEIARKYIDALGGRENVNSVAHCATRLRVKDKVDVDRQSLDDPDLKGTFFETGGM 60  
QY 61 FOITVGPQDVPDFEKE---LDDATSKDIATVSTEQLKQDVANNANWFSRAVKVLADIPVPL 117

Db 61 YQIFGTGTNNMYDEVVVLGLPTS-----SKDMKAEVAQGNWFQRAIRTEGDFVFPV 115  
QY 118 IPILVGGGLMANNVLAODLFGPOSIVEMFQISGVAEMINLMASAPFAFLPVLVGFT 177  
Db 116 IPVIVATGLFWGVRGLENALXM-----PLPGDEATYITQILTDTAIFILPGLVWVS 165  
QY 178 ATRKFGNEFLGAGIGMAMVFPVLVNGYDVAATAGEMPMSLFLGLDVAQAGYQGTVLP 237  
Db 166 TFRVFGNRPVIGIVLGMMLVSGSLPNAWAVA---QGGEVTAMNFFGP-IPVGLQSGSLP 221  
QY 238 VLVSWSLITATEKFLHKLMTADFLITPVLTLTLGLTFLFIAIGPAMRWGDLIAHGLQ 297  
Db 222 AFILGVVGAKEKAVKRVDPDVIDLVTPTVLLVXLSILGLFVGPVHVVENVILIATK 281  
QY 298 GLYDFGPGVGLLGLVYSPIVITGLHOSPPTELELFLNQGGSFIF-ATASMANIAQGA 356  
Db 282 AILSNPFLGGLGFLGGVHQIIVSVGVHIFNLLEVQLLAADHANPFNAITAAATQGA 341  
QY 357 CLAYFFLAKSKLGLAGAGSVAVLGITBPAIFGNLRLRWPFFYIGIGTAAIGGALIAL 416  
Db 342 TVAVGVTKNPKLTALFAPALSAFLGITPAIFGNLRLRWPFFYIGIGTAAIGGALIAL 401  
QY 417 FDIKAVALGAAGFLGVVYSIDAPDV---MFLYCAVVTFTVAFGAATAY-----GLYIV 468  
Db 402 -----LGLAGTNGTITIPGTMLYVNGOLPOYLMLVAVSFALGFALTMYMEFY 449  
QY 469 NSGIDDPATAAPVPAGTAKAE-APAEFSNDSTIIQAPLTGEAIALSSVSDFAMFAGK 527  
Db 450 --EDEVDATAAKAEVAEKEEVAAPALQNETIV--TPIVGDVVVALDNDVPFSSGAM 505  
QY 528 GSGVAIVTKGOLVSPVSGKIIVAPPSGHAFAVRTKAEDGNSVDILMHIGDVTYNLNGT 587  
Db 506 GOGIATVKSQGVVYALADAESAIAFPTGHAFGLKTR-----NGAEVLIVHIGDVTYVMDG 561  
QY 588 FNPLKKQDEVKAGELLCEFDIDAIAKAGYEVTTPIVYSN---YKKTGPNVTYGLGTEA 644  
Db 562 FEAKVAQGNKVRKAGDVLGTGFSNKAIAAGLDDTTMVIIVTADVASVAPVAT---GSVSK 618  
QY 645 GANLLNV 651  
Db 619 GDAVIEV 625

RESULT 14  
C95220  
Trehalose PTS system, IIABC components [Imported] - Streptococcus pneumoniae (strain  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: C95220  
R:Petelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: C95220  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-655 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75956.1; PID:g14973388; GSPDB:GN00164; TIGR:  
A:Experimental source: strain TIGR4  
C:Genetics:  
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera  
A:Gene: Spt84  
Query Match 28.4%; Score 949; DB 2; Length 655;  
Best Local Similarity 34.9%; Pred. No. 1.1e-56;  
Matches 234; Conservative 128; Mismatches 253; Indels 56; Gaps 19;  
QY 7 AQRILRDIGEDNIVAAACATRLRLVLKDKVDVDRQSLDDPDLKGTFFETGGMFOIIVG 66  
Db 8 AKDLQAGGKENVATVTHCATRMRFLVGLDDKKNVKAIESIPAVKGTFTNAGQFQVILG 67



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 01:26:49 ; Search time 24 Seconds  
(without alignments)  
1295.194 Million cell updates/sec

Title: US-10-019-284A-2

Perfect score: 3342

Sequence: 1 MDHKDLAQRLTRDIGGEDNI.....IEAGNLLNVAKKEAVPATP 661

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1609.5	48.2	651	1	PTSA_PEDPE
2	1478.5	44.2	664	1	PTSA_STRMU
3	1277.5	36.7	480	1	PTSB_STAXY
4	1100.3	32.9	474	1	PTSB_PASMU
5	1069.3	32.0	479	1	PTSB_VIBAL
6	999.3	29.9	470	1	PTSB_BACSU
7	901.3	27.0	609	1	PTBA_BACSU
8	896.3	26.8	631	1	PTBA_ERMCH
9	824.3	24.7	625	1	PTBA_ECOLI
10	819.3	24.5	456	1	PTSB_SALTY
11	816.3	24.4	456	1	PTSB_KLEPN
12	751.5	22.5	460	1	PTSB_BACSU
13	725.5	21.7	683	1	PTNA_CORGL
14	712.5	21.3	674	1	PTGA_CORGL
15	706.3	21.1	459	1	SACX_BACSU
16	655.3	19.6	473	1	PTTB_ECOLI
17	438.3	13.1	474	1	YFEV_ECOLI
18	437.3	13.1	482	1	PTSB_VIBCH
19	436.5	13.1	485	1	PTDA_ECOLI
20	353.3	10.6	699	1	PTGA_BACSU
21	329.3	9.8	651	1	PTAA_KLEPN
22	314.5	9.4	726	1	PTGA_STRPN
23	313.5	9.4	631	1	YFBS_BACSU
24	308.5	9.2	648	1	PTAA_ECOLI
25	299.5	9.0	324	1	PTGA_BACST
26	286.3	8.6	634	1	LACY_STRTR
27	256.3	7.7	189	1	PTGA_BORBU
28	255.3	7.6	168	1	YFOE_BACSU
29	255.3	7.5	627	1	LACY_LACDE
30	249.5	7.5	154	1	PTGA_MYCCA
31	246.3	7.4	168	1	PTGA_ECOLI
32	244.3	7.3	168	1	PTGA_SALTY
33	240.5	7.2	641	1	RAPP_PEDPE

34	238.5	7.1	165	1	PTGA_HAEIN	P45338 haemophilus
35	231.5	6.9	161	1	PTGA_BUCAL	O9xii7 buchnera ap
36	228.5	6.8	940	1	PTGA_MYCPN	P75569 mycoplasma
37	225.5	6.7	167	1	PTGA_BUCBP	Q8ka51 buchnera ap
38	212.3	6.3	168	1	PTGA_MYCGE	O89b05 buchnera ap
39	208.5	6.2	908	1	PTGA_MYCGE	P47315 mycoplasma
40	205.5	6.1	483	1	PTVB_ECOLI	P32154 escherichia
41	185.5	5.6	578	1	PTFB_RHOCA	P23387 rhodobacter
42	172.5	5.2	658	1	HRSA_ECOLI	P54745 escherichia
43	169.5	5.1	639	1	LACY_LEULA	O48624 leuconostoc
44	163.5	4.9	359	1	PTWC_ECOLI	P36672 escherichia
45	162.3	4.8	527	1	PTTB_BACSU	P54715 bacillus su

ALIGNMENTS

RESULT 1

PTSA_PEDPE	STANDARD;	PRT;	651 AA.
ID	PTSA_PEDPE		
AC	P43470;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	PTS system, sucrose-specific IIAIC component (EIIABC-SCR) (Sucrose-		
DE	permease IIAIC component) (Phosphotransferase enzyme II, ABC		
DE	component) (EC 2.7.1.69) (EII-SCR).		
GN	SCRA.		
OS	Pediococcus pentosaceus.		
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pediococcus.		
OX	NCBI_TaxID=1255;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PEP1.0;		
RA	Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;		
RL	Submitted (xxx-1994) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT		
CC	SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE		
CC	-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE		
CC	AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY		
CC	PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS		
CC	PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO		
CC	THE SUGAR.		
CC	-1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein		
CC	histidine + sugar phosphate.		
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.		
CC	-1- SIMILARITY: Contains 1 PTS EIIA domain.		
CC	-1- SIMILARITY: Contains 1 PTS EIIB domain.		
CC	-1- SIMILARITY: Contains 1 PTS EIIC domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; Z32771; CAA83668.1; -		
DR	EMBL; L32093; AAA25567.1; -		
DR	PIR; S44257; S44257.		
DR	HSP; P20166; IGPR.		
DR	InterPro; IPR001127; PTS_EIIA.		
DR	InterPro; IPR001996; PTS_EIIB.		
DR	InterPro; IPR003352; PTS_EIIC.		
DR	Pfam; PF00358; PTS_EIIA_1.		
DR	Pfam; PF00367; PTS_EIIB_1.		
DR	Pfam; PF02378; PTS_EIIC_1.		
DR	ProDom; PD002243; PTS_EIIA; 1.		
DR	ProDom; PD001476; PTS_EIIB; 1.		
DR	TIGRFAMS; TIGR00826; EIIB_glc; 1.		
DR	TIGRFAMS; TIGR00830; PTBA; 1.		
DR	PROSITE; PS00371; PTS_EIIA_1; 1.		

RESULT 2	PTSA_STR	PTS	P12	01-	01-	28-	PTS	per
	ID	AC	DT	DT	DT	DE	DE	

Phosphorylation; Transmembrane; Complete proteome.

FT DOMAIN 1 40 EIIB DOMAIN.  
 FT DOMAIN ? ? EIIB DOMAIN.  
 FT DOMAIN 533 664 EIIB DOMAIN.  
 FT MOD\_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 664 AA; 69988 MW; 809P63E32281A9A1 CRC64;

Query Match 44.2%; Score 1478.5; DB 1; Length 664;  
 Best Local Similarity 46.0%; Pred. No. 1.9e-82;  
 Matches 311; Conservative 127; Mismatches 203; Indels 35; Gaps 13;

QY 1 MDHKDLAQRILRDIGGEDNIVAAAHCAHRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60  
 DB 1 MDXKVAASEVITAV-GKDNLVAAAHCAHRLRLVLKDDSKVDQKALDKNADVAGTKTDQ 59  
 QY 61 FOIIVPGDGDVHVFKELDATSKDIAVSTEQKLDVVANNA--NWFSRAVKVLADIFVPLI 118  
 DB 60 YQVIIGPGDGVNFYDIIRKQTL-TEVSTDDKTKAAAGKFPNIMALKLLSDIFVPII 118  
 QY 119 PILVGGGLLMAINNVVLAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 178  
 DB 119 PALVAGGLLMAINNVVLAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 178  
 QY 179 TKRFGNEFLGAGIGMANVFPILVNGYVAA-----TWYAGEMP-MWSEFLGLDVAQAGY 231  
 DB 179 AKRFGANQFLGASIGMIVAPGAANTIGLAANAPISKAATIGAYTGFNIFGLHVTQASY 238  
 QY 232 QGVPLVPLVYVSWLTAIEFKRLKMGTAFLITPVLTLGLLGFLLTFFIAGFAMRWVGL 291  
 DB 239 TYQVIVPLVAVMLLSILEFFKRLPSAVDFTFPLLSVLIIGLTFIIVGPMKEVSDM 298  
 QY 292 LAHGLQGLDFGPGVGLLFLGLVYSPVITGLHQSPPILEL---FNQG---GSEFIFAT 345  
 DB 299 LTRGIYVWLYDTTGLGNGVFGALYSPVMTGLHQSPPILEL---FNQG---GSEFIFAT 358  
 QY 346 ASMANAQAACLAFFFLAKSEKLGAGAGSVAVLGTETPAIFGVNLRWLPFFVIGIG 405  
 DB 359 ASMANAQAATTAIFLTKDKKGLSSGVSALLGTEPAIFGVNLRWLPFFVIGIG 418  
 QY 406 TAIGGALIALFDIKVAALGAAGFLGWSIDAPDMVFLVCVAVFVIAFGAAIAGLYL 465  
 DB 419 GSASAAIAAGLLQVAVSLGSAFLGLSTKASSIPFYVVCELISPAIAFAVTVYGV--- 475  
 QY 466 VRRNGSDPDATAPVAGTAKAEAPAEFSNDST-----LIQAPLTGEAIALSSVS 518  
 DB 476 --KTKAVDVFPAEAAREAEAEVQ-EIPEAASAAKAAQVDEVLAPLAGAEVETSYN 532  
 QY 519 DAMFASGLGSGVAIVPTKQLVSPVSGKIVVAFPSGHAFAYKTAEDGNSVDILMHIGF 578  
 DB 533 DPVFSSEAMKGIAIKPSGNTVYAPVDGTQIAFDTHAYGI--KSDNGA--EILIHIGI 588  
 QY 579 DTVNLNGTHFNPLKQGDVYKAGELLCEDIDAIAKAGYEVTTPIVVSNNYKKTGPVNTY- 637  
 DB 589 DTVSMGEGFEQVQADQIKKGDVLGTFDSKIAEAGLNTMTFIVTADYASVETLA 648  
 QY 638 GLGEIEAGANLLNVAK 653  
 DB 649 SSGTAVAGDSILEVKK 664

RESULT 3  
 PTSEB\_STAXY STANDARD; PRT; 480 AA.  
 AC P51184;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIB-SCR) (Sucrose-  
 permease IIBC component) (Phosphotransferase enzyme II, BC component)  
 DE (EC 2.7.1.69) (EIIB-SCR).  
 GN SCRA.

OS Staphylococcus xylosus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_taxid=1288;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=DSM 20267 / Isolate C2A;  
 RC MEDLINE=94049686; PubMed=8232209;  
 RA Wagner E., Goetz F., Brueckner R.;  
 RX "Cloning and characterization of the scrA gene encoding the sucrose-  
 specific Enzyme II of the phosphotransferase system from  
 Staphylococcus xylosus";  
 RL Mol. Gen. Genet. 241:33-41(1993).  
 CC !- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 SUGAR PHOTOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE  
 AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPP); IIA TRANSFERS ITS  
 PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 THE SUGAR.  
 CC !- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 histidine + sugar phosphate.  
 CC !- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC !- SIMILARITY: Contains 1 PTS EIIB domain.  
 CC !- SIMILARITY: Contains 1 PTS EIIC domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X69800; CA49461.1; .  
 DR PIR; S39978; S39978.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; P0001476; PTS\_EIIB; 1.  
 DR TIGRfam; TIGR00826; EIIB\_9lc; 1.  
 DR PROSITE; PS01035; PTS\_EIIB-CYS; 1.  
 DR Phosphotransferase system; Sugar transport; Transferase;  
 KW Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 ? EIIB DOMAIN.  
 FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 480 AA; 51326 MW; AB4ED9765D84E47 CRC64;

Query Match 36.7%; Score 1227; DB 1; Length 480;  
 Best Local Similarity 52.7%; Pred. No. 2.3e-67;  
 Matches 252; Conservative 83; Mismatches 133; Indels 10; Gaps 6;

QY 1 MDHKDLAQRILRDIGGEDNIVAAAHCAHRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60  
 DB 1 MDXKVAASEVITAV-GKDNLVAAAHCAHRLRLVLKDDSKVDQKALDKNADVAGTKTDQ 59  
 QY 61 FOIIVPGDGDVHVFKELDATSKDIAVSTEQKLDVVANNA--NWFSRAVKVLADIFVPLI 120  
 DB 60 YQVIIGPGDGVNFYDIIRKQTL-TEVSTDDKTKAAAGKFPNIMALKLLSDIFVPII 119  
 QY 121 LVGGGLLMAINNVVLAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 179  
 DB 120 IVAGGLLMAINNVVLAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 179  
 QY 180 KRFGGNEFLGAGIGMANVFPILVNGYVAA--EMPMWSLFLGLDVAQAGYQGTVPV 238  
 DB 180 KRFGGNAYLGAALGMLVHPPELMSAYDYPKALEAGKEPHNLFGLINQVGYQGVLP 239  
 QY 239 LVVSWILATIEFKRLKMGTAFLITPVLTLGLLGFLLTFFIAGFAMRWVGLDLAGLQ 298  
 DB 240 LVATYILATIEFKRLKMGTAFLITPVLTLGLLGFLLTFFIAGFAMRWVGLDLAGLQ 299





[illegible]

"Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the hut and wpa loci.",  
 Microbiology 141:337-343(1995).  
 [3]  
 RC STRAIN-168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Meszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly J.E., Grandi G.,  
 RA Guseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kashahara Y., Klearr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maubel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prascan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche B., Rose M., Sadale F.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognori A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RL Nature 390:249-256(1997).  
 [4]  
 RN SEQUENCE OF 1-192 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE=96204517; PubMed=8628237;  
 RA Beloin C., Hirschbein L., le Hegarat F.,  
 RT "Suppression of the Bgl+ phenotype of a delta hns strain of  
 RT *Escherichia coli* by a *Bacillus subtilis* antiterminator binding  
 RT site,"  
 RL Mol. Gen. Genet. 250:761-766(1996).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 CC histidine + sugar phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Contains 1 PTS EIIA domain.  
 CC -!- SIMILARITY: Contains 1 PTS EIIB domain.  
 CC -!- SIMILARITY: Contains 1 PTS EIIC domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; 234526; CA84286.1; -  
 DR EMBL; D31856; BAA06652.1; -  
 DR EMBL; D29985; BAA06256.1; -  
 DR EMBL; 299124; CAB15963.1; -

DR EMBL; X85408; CAA59697.1; -  
 DR PIR; I40406; I40406.  
 DR PIR; T47097; T47097.  
 DR HSSP; P20166; LGPR.  
 DR SUBTILIST; BG10934; bg1p.  
 DR InterPro; IPR001127; PTS\_EIIA.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00358; PTS\_EIIA.1; 1.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD002243; PTS\_EIIA; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR TIGRFAMS; TIGR00830; PTBA; 1.  
 DR PROSITE; PS00371; PTS\_EIIA.1; 1.  
 DR PROSITE; PS01035; PTS\_EIIB-Cys; 1.  
 KW Phosphotransferase system; Sugar transport; Transferase;  
 KW Phosphorylation; Transmembrane; Complete proteome.  
 FT DOMAIN 1 41 EIIB DOMAIN.  
 FT DOMAIN ? ? EIIC DOMAIN.  
 FT DOMAIN 480 609 EIIA DOMAIN.  
 FT MOD\_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 302 302 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 532 532 PHOSPHORYLATION (BY SIMILARITY).  
 FT TRANSMEM 112 132 POTENTIAL.  
 FT TRANSMEM 141 161 POTENTIAL.  
 FT TRANSMEM 174 194 POTENTIAL.  
 FT TRANSMEM 202 222 POTENTIAL.  
 FT TRANSMEM 246 266 POTENTIAL.  
 FT TRANSMEM 281 301 POTENTIAL.  
 FT TRANSMEM 321 341 POTENTIAL.  
 FT TRANSMEM 351 371 POTENTIAL.  
 FT TRANSMEM 379 399 POTENTIAL.  
 FT TRANSMEM 412 432 POTENTIAL.  
 FT CONFLICT 75 75 A -> S (IN REF. 4).  
 FT CONFLICT 288 288 L -> F (IN REF. 2).  
 FT CONFLICT 434 434 E -> G (IN REF. 2).  
 FT CONFLICT 436 436 A -> S (IN REF. 2).  
 FT CONFLICT 449 450 DG -> HR (IN REF. 2).  
 FT CONFLICT 549 549 I -> M (IN REF. 2).  
 FT CONFLICT 552 552 G -> S (IN REF. 2).  
 SQ SEQUENCE 609 AA; 64550 MW; 5F630C671D21FBED CRC64;  
 Query Match 27.0%; Score 901; DB 1; Length 609;  
 Best Local Similarity 32.3%; Pred. No. 1.6e-47;  
 Matches 215; Conservative 127; Mismatches 258; Indels 66; Gaps 14;  
 Qy 1 MDHKDLAQRILRDIGGEDNIVAAAHACATRLRLVVKDKVDVDRSLDDDDPKLGTFTTGGM 60  
 Db 1 MDYDKLSKDLQLVGGGEENVQVIRHCHMTRLRFLNHDNAKADRSQSLQSLQSLQSLQSLQSLQ 60  
 Qy 61 FQITVGGPDVHDVFKEL---DDATSKDIAVSTQLKDVVANNANWFSRAVKVLADIFVPL 117  
 Db 61 FQITVGGPDVHDVFKEL---DDATSKDIAVSTQLKDVVANNANWFSRAVKVLADIFVPL 117  
 Qy 61 FQITVGGPDVHDVFKEL---DDATSKDIAVSTQLKDVVANNANWFSRAVKVLADIFVPL 117  
 Db 61 FQITVGGPDVHDVFKEL---DDATSKDIAVSTQLKDVVANNANWFSRAVKVLADIFVPL 117  
 Qy 118 IPIVGGGLMALNNVLVAQDLFGPSLVEMFFQISGVAEMINLMASAPAFPLVVGFT 177  
 Db 118 IPIVGGGLMALNNVLVAQDLFGPSLVEMFFQISGVAEMINLMASAPAFPLVVGFT 177  
 Qy 113 LPAIAGAGMIKGLVALAVT---EG-----WMAEKSQVHVILTAVDGAGFYFPLDLAMS 163  
 Db 113 LPAIAGAGMIKGLVALAVT---EG-----WMAEKSQVHVILTAVDGAGFYFPLDLAMS 163  
 Qy 178 ATRFRGGNEFLGAGIGAMVFTLVNGYDVAATAGEMPMSLSFLGLDVAQAQYQGTPLP 237  
 Db 178 ATRFRGGNEFLGAGIGAMVFTLVNGYDVAATAGEMPMSLSFLGLDVAQAQYQGTPLP 237  
 Qy 164 AARFGSNPVAAIAAAILHP-----DLTALLGAGK-PI-SFGLPVTAATYSSTVIP 215  
 Db 164 AARFGSNPVAAIAAAILHP-----DLTALLGAGK-PI-SFGLPVTAATYSSTVIP 215  
 Qy 238 VLVVSWTLATIEKFLKRLMGTAFLITPVLTLLGLFTLFTLFTLFTLFTLFTLFTLFTLFTLFTL 297  
 Db 238 VLVVSWTLATIEKFLKRLMGTAFLITPVLTLLGLFTLFTLFTLFTLFTLFTLFTLFTLFTLFTL 297  
 Qy 216 ILLISNTASVKEKWLDFTHASLKLIVVPTFTLITVPLTITVPLTITVPLTITVPLTITVPLTITV 275  
 Db 216 ILLISNTASVKEKWLDFTHASLKLIVVPTFTLITVPLTITVPLTITVPLTITVPLTITVPLTITV 275  
 Qy 298 GLVDFGPGVGLLFGVLVSPVITGLHSPFPIELFNQGG-SFIFATASMANIAQGA 356  
 Db 298 GLVDFGPGVGLLFGVLVSPVITGLHSPFPIELFNQGG-SFIFATASMANIAQGA 356  
 Qy 276 YLFDHAGLVAMILLAGTFSLIIMTGHYAFVPMINNIAGNGHDYLLPAMFLANMQAGA 335  
 Db 276 YLFDHAGLVAMILLAGTFSLIIMTGHYAFVPMINNIAGNGHDYLLPAMFLANMQAGA 335  
 Qy 357 CLAVFFLAKSEKLKGLAGSGVSAVLGITPEIPATFGVNLRLRWPFFYIGITGTAAGGALIAL 416  
 Db 357 CLAVFFLAKSEKLKGLAGSGVSAVLGITPEIPATFGVNLRLRWPFFYIGITGTAAGGALIAL 416

```

Db 336 SPAVFLSRNKKFKSLALTSITLALMGITEPAMVGNMRLKPPFAALIGGAAGAFYGM 395
QY 417 FDIKAVALGA-AGFLGVWSIDAPDMVFLVCAVTFVIAFGAAIAYGLVLRNGSIDPD 475
Db 396 TGVASIVGNGAGLPSIPFVIGFTFIYAFIAFAETAATAAALLGFDVPSDGS---- 451
QY 476 ATAAPVAGCTTKAEAPAEAFESNDSTIIIOAPLTGEAIALSSVSDAMFASGLSGVAIVP 535
Db 452 -----QQPAVHEGSRRIIHSPIKGEVKALSEVKGVSAGVMKGFALBP 496
QY 536 TKQLVSPVSGKIIVAFPPSGHAFVARTKAEDGSNDVILMHIGFDTVNLNGTFFNPLKKOG 595
Db 497 ECEGVSPVSGVSTTFIKTKHAIGITS----DQGAELIHIGLDTVKLEQSQFTHIKEG 552
QY 596 DEYKAGELCEFDIDAIRKAGVEVTPVIVSNYKK--TGPVNTYGLGEIEAGANLLNAYK 653
Db 553 DRVAPGDPLVLEQIKRAGYDVIIPVIVTNDQYSFSPVKEIG-----KVOP 601
QY 654 KEAPPA 659
Db 602 KEALLA 607

RESULT 8
PTBA_ERWCH STANDARD; PRT; 531 AA.
AC P26207;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PIS system, beta-glucoside-specific IIBC component (EIIABC-Bgl)
DE (beta-glucoside-permease IIBC component) (Phosphotransferase
DE enzyme II, ABC component) (SC 2.7.1.69) (EII-Bgl).
GN ARBG.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92121114; PubMed=1732212;
RX el Hassouni M., Henrissat B., Chippaux M., Barras F.;
RT "Nucleotide sequences of the arb genes, which control beta-glucoside
RT utilization in Erwinia chrysanthemi: comparison with the Escherichia
RT coli bgl operon and evidence for a new beta-glycohydrolase family
RT including enzymes from eubacteria, archaeobacteria, and humans.";
RL J. Bacteriol. 174:765-777(1992).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON ARBG (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: Contains 1 PTS EIIA domain.
CC -1- SIMILARITY: Contains 1 PTS EIIIB domain.
CC -1- SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL: M81772; AAA24814.1;
DR PIR: B42603; B42603.

```

```

DR HSP; P20166; IAX3.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA.1; 1.
DR Pfam; PF00367; PTS_EIIB.1.
DR Pfam; PF02378; PTS_EIIC.1.
DR ProDom; PD002443; PTS_EIIA.1.
DR ProDom; PD001476; PTS_EIIB.1.
DR TIGRFAMs; TIGR00830; PTBA; 1.
DR PROSITE; PS00371; PTS_EIIA.1; 1.
DR PROSITE; PS00371; PTS_EIIB.CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Inner membrane.
FT DOMAIN 1 41
FT DOMAIN ? 41
FT DOMAIN 501 631
FT MOD_RES 26 26
FT MOD_RES 309 309
FT MOD_RES 553 553
FT TRANSMEM 50 70
FT TRANSMEM 99 119
FT TRANSMEM 120 140
FT TRANSMEM 146 166
FT TRANSMEM 175 195
FT TRANSMEM 206 226
FT TRANSMEM 248 268
FT TRANSMEM 272 292
FT TRANSMEM 295 315
FT TRANSMEM 328 348
FT TRANSMEM 358 378
FT TRANSMEM 385 405
FT TRANSMEM 407 427
FT TRANSMEM 434 454
SQ SEQUENCE 631 AA; 66984 MW; 7D0BD27A36BFFDF CRC64;

Query Match 26.8%; Score 896; DB 1; Length 631;
Best Local Similarity 32.7%; Pred. No. 3.4e-47;
Matches 214; Conservative 128; Mismatches 278; Indels 34; Gaps 12;

QY 1 MDKDLAQRILRDIGGEDNIVAAAHCAFLRLVLKDTKVDQRQSLDDDDPDLLKGFETGGM 60
Db 1 MNYETLASEIRDGVGGOENIIVHICATFLRFLKRLDNTINADALKNNPGIIMVVEGQ 60
QY 61 FOIIVG--PGVDVHVFKELDD-ATSKDIAVSTEQLKDVVANNANWFSAVKLADIFVPL 117
Db 61 FQVVVGNQVADVYQALLSLDGMARFSDSAPEEKKN-----SLFSGFDIISISITPF 114
QY 118 IPILVGGGLLMAINNVLAQDLFGPSQSLVEMFPQISGVAEMINLMASAPFAFLPVLYGFT 177
Db 115 VGVMAATGILGFLALGVA-----THVISESGTYKLLFAASDALFFFFVLGYT 165
QY 178 ATKRFPGNEFLGAGIGMAWVFTLVNGVDVAATWTAGEMPWSLFGLDVAOAGYGVLP 237
Db 166 AGKRFGGNFTTLVIGATLVHPSMTAFN---AMQAPDHSTLHFLGIPITINYSVYP 222
QY 238 VLVVSWILATIEKFLHKLKRLMGTAFLITPVLTLTLTGFTFIATGAPMRWVGLLHGLQ 297
Db 223 ILFASWVCKLEKPLNRLWLNHNIRNFTPLLCIVISVPLTELLGPNATWLSQMLAGGY 282
QY 298 GLYDFGPGVGLLFGLVYSPVITGLHQSFPPPIELFNQ--GGSFIFATASMANIAOGAA 356
Db 283 WLYGLNSLLAGAVMGALWQVCVIFGLHWGFVPLMNNFSVIGHDTLLPLLVPAVLGQAGA 342
QY 357 CLAVFFLAKSKLKGAGSGVSAVLGITEPAIGVNLRLRWYPIGIGTAAGALIAL 416
Db 343 TLGVLLRTQDLKRKGIAGSAFSAIFGITEPAYGVTLPLRRPPIFOCIGALGAANGY 402
QY 417 FDIKAVALGAAGFLGVWSIDAPDMVFLV-CAVTFVIAFG-AAIAYGLVLRNGSIDP 474
Db 403 AHTMYSGFSISFTQVIPTGVDSVSWAAVIGTLLAFAPALTSVSGVPKD---ET 459
QY 475 DATAAPVAGCTTKAEAPAEAFESNDSTIIIOAPLTGEAIALSSVSDAMFASGLSGVAIV 534

```





Qy 181 RFGNBEFLGAGIGAMVFTLVNGYDVA--TMTAGEMPMSLFGLDVAQAGYQGTVP 237  
 Db 171 EFGNPLYGATLGGIILTHPALTNAMGVAAGFTM-----NFFGEFAMIGYGTVP 222  
 Qy 238 VLVSWSLAFIEFLHKLRLMGTAFLITPVLTLTLGFLTFIAIGPAMRWGDLAHLGQ 297  
 Db 223 VLLAVWFMISVEKQLRAIPDALDLILTFPLTVIISGFALLIIPAGRALGDGIFVLS 282  
 Qy 298 GLYDFGPGVGLLFGVLVYSPVITGLHQSPFPPIELELFNQ---GGSFIFATASMANIAOG 354  
 Db 283 TLISHAGWLAGLFGGLYVIVITGHSFHAVERAGLLGNPSIGVNFLLPIWAMANVAOG 342  
 Qy 355 AACLAVFELAKSEKLAGAGSVAVLGITTEPAIFGNVLRWRPFIYIGITAAIGGALI 414  
 Db 343 GACFVFWKTKDKAKITAILPSAFSAMLGITEAIFGINLRFVKPFIALVGAAGAVW 402  
 Qy 415 ALFDIKAVAGAGFLGVSVISDAPDMWFMELVCVAVTFVAFGAIAIYL 465  
 Db 403 VSVHVMYMTAVGLTAIPGMAIVQASSLLNY----IIGWIAFGVAFVSLVL 449

## RESULT 11

PTSB\_KLEPN STANDARD; PRT; 456 AA.  
 AC P27219;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-  
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)  
 DE (EC 2.7.1.69) (EII-SCR).  
 GN SCRA.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1033-5p14 / KAY2026;  
 RX MEDLINE=9618840; PubMed=8628219;  
 RA Tigemeyer F., Jahreis K., Ebner R., Lengeler J.W.:  
 RT "Molecular analysis of the scrA and scrB genes from Klebsiella  
 RT pneumoniae and plasmid pUR400, which encode the sucrose transport  
 RT protein Enzyme II scr of the phosphotransferase system and a  
 RT sucrose-6-phosphate invertase".  
 RL Mol. Gen. Genet. 250:197-206(1996).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A  
 CC COVALENTLY BOUND EIIB DOMAIN. INSTEAD, EII-SCR-MEDIATED  
 CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-  
 CC GIC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.  
 CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 CC histidine + sugar phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -!- SIMILARITY: Contains 1 PTS EIIB domain.  
 CC -!- SIMILARITY: Contains 1 PTS EIIC domain.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X57401; CAA40658.1; .  
 DR PIR; S62331; S62331.

DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR InterPro; IPR004719; PTS\_EIIC\_gic.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR TIGRFAMS; TIGR00826; EIIB\_gic; 1.  
 DR TIGRFAMS; TIGR00852; pts\_gic; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Transport; Sugar transport; Transferase;  
 KW Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 111 EIIB DOMAIN.  
 FT DOMAIN 112 456 EIIC DOMAIN.  
 FT TRANSMEM 112 132 POTENTIAL.  
 FT TRANSMEM 144 164 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT TRANSMEM 247 267 POTENTIAL.  
 FT TRANSMEM 288 308 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 FT TRANSMEM 360 380 POTENTIAL.  
 FT TRANSMEM 388 408 POTENTIAL.  
 FT TRANSMEM 428 448 POTENTIAL.  
 FT MOD\_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 456 AA; 48022 MW; 4AFDF5405CAEFC66 CRC64;  
 Query Match 24.4%; Score 816; DB 1; Length 456;  
 Best Local Similarity 37.8%; Pred. No. 1.7e-42;  
 Matches 176; Conservative 88; Mismatches 177; Indels 24; Gaps 5;  
 Qy 1 MDKDLAQRILRDIGGEDNIVAAHCAHRLVLRKTDKVDQRSLDDDDPLKGTETGGM 60  
 Db 1 MDEFQISRLPLLGKENIASAAHCAHRLVLRVDDALDAQAIGKIDGVKGCERNAGQ 60  
 Qy 61 FOIIVGPGDVHVFKELDLDDATSKDIAVSTEQLKDVANNANWFSRAVKVLADIFVPLPI 120  
 Db 61 MQIIFGTGVNKYAFAFIOQAIGISE-SKSEAADAAKKNLPFQRIARLLSNIFPIIPA 119  
 Qy 121 LVGGGLLMAINNVLAQDLFGPQSLVEMPPQISGVAEMINLMASAPFAFLPVLVGFATK 180  
 Db 120 IVASGLLGLLGWVKTYGVDPNSALYI-----MLDMCSSAAFIILPIIGFTAAR 170  
 Qy 181 REGNFEFLGAGIGAMVFTLVNGYDVA--TMTAGEMPMSLFGLDVAQAGYQGTVP 237  
 Db 171 EFGNPLYGATLGGIILTHPALTNAMGVAAGFTM-----NFFGEFAMIGYGTVP 222  
 Qy 238 VLVSWSLAFIEFLHKLRLMGTAFLITPVLTLTLGFLTFIAIGPAMRWGDLAHLGQ 297  
 Db 223 VLLAVWFMISVEKQLRAIPDALDLILTFPLTVIISGFALLIIPAGRALGDGIFVLS 282  
 Qy 298 GLYDFGPGVGLLFGVLVYSPVITGLHQSPFPPIELELFNQ---GGSFIFATASMANIAOG 354  
 Db 283 TLISHAGWLAGLFGGLYVIVITGHSFHAVERAGLLGNPSIGVNFLLPIWAMANVAOG 342  
 Qy 355 AACLAVFELAKSEKLAGAGSVAVLGITTEPAIFGNVLRWRPFIYIGITAAIGGALI 414  
 Db 343 GACFVFWKTKDKAKITAILPSAFSAMLGITEAIFGINLRFVKPFIALVGAAGAVW 402  
 Qy 415 ALFDIKAVAGAGFLGVSVISDAPDMWFMELVCVAVTFVAFGAIAI 459  
 Db 403 VSVHVMYMTAVGLTAIPGMAIVQASSLLNYIIGMAIAFAVAFALSL 447

## RESULT 12

PTSB\_BACSU STANDARD; PRT; 460 AA.  
 ID PTSB\_BACSU  
 AC P05306;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-  
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)  
 DE (EC 2.7.1.69) (EII-SCR).



```

Db 343 GLAVFWAKAKTKEALPAASAFGLCIETPEVIFGVNLYRKPFFIAAMIGGALGGAYVVF 402
QY 417 FDIKAVAGAGFLGVVSDAP-----DMVNFVCAVTVFVIFAGAA 458
Db 403 THVAANAYGLTG-IPMIAAAPFGFNSNLHYLIGMAIAAASAFIAA 447

RESULT 13
PTNA_CORGL
ID PTNA_CORGL STANDARD; PRT; 683 AA.
AC Q45072;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PFS system, mannose-specific IIABC component (EIIABC-Man) (Mannose-
DE permease IIABC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.65) (EII-Man/EIIC-Man).
GN PTSM OR CGL1360.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94314161; PubMed=8039653;
RA Lee J.K., Sung M.H., Yoon K.H., Yu J.H., Oh T.K.;
RA "Nucleotide sequence of the gene encoding the Corynebacterium
RA glutamicum mannose enzyme II and analyses of the deduced protein
RA sequence.";
RL FEMS Microbiol. Lett. 119:137-145(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar -> protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Contains 1 PTS EIIA domain.
CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L18974; AAA53546.1; .
CC EMBL; AP005278; BAB98753.1; .
CC HSSP; P08837; 1GIC.
CC InterPro; IPR001127; PTS-EIIA.
CC InterPro; IPR001596; PTS-EIIB.
CC InterPro; IPR003352; PTS-EIIC.
CC Pfam; PF00358; PTS-EIIA_1; 1.
CC Pfam; PF00367; PTS-EIIB; 1.
CC Pfam; PF02378; PTS-EIIC; 1.
CC ProDom; PD002243; PTS-EIIA; 1.
CC ProDom; PD001476; PTS-EIIB; 1.
CC TIGRFAMs; TIGR00830; PTBA; 1.
CC PROSITE; PS00371; PTS-EIIA_1; 1.
CC PROSITE; PS01035; PTS-EIIB_CYS; 1.

```

```

KW Phosphotransferase system: Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Complete proteome.
FT DOMAIN 1 43 EIIA DOMAIN.
FT ? 683 EIIA DOMAIN.
FT DOMAIN 126 146 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
FT MOD_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 602 602 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 683 AA; 72571 MW; 0B42CAREC60828075 CRC64;

Query Match 21.7%; Score 725.5; DB 1; Length 683;
Best Local Similarity 30.9%; Pred. No. 8e-37;
Matches 216; Conservative 119; Mismatches 286; Indels 79; Gaps 18;

QY 7 AQRILRDIGGEDNIVAAAHACATRLRLVKDKDVRQSLDDDDPDLKGTFTG--GMFOII 64
Db 9 SCHILENUGGPNITSMTHCATRLRFQVKDQISIVDQOEIDSPSVLGVVPGSTGM-QVV 67
QY 65 VFGPDVDHVFKEI-----DQATSKDIADVSTEQKQKOVYANNANWFSRAVKVLADI 113
Db 68 MG-GSVANYVYQILKLDGKMKHFADEAVES---SSKKEYGVGRKYSWIDVAFELSDT 122
QY 114 FVPLPILVGGGLLMAINNVLVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVL 173
Db 123 FRPILWALLGSLIITL---LVLDATFGLDQDFRPMDEQPDITYVFLHSMWSVFFLPIM 179
QY 174 VGFTATKRGGEFLGAGIGAMKVPFTLVNGVDVAATMTAGEMPMSFLGLDVAQAGYQG 233
Db 180 VGATAARKLGANEWIGAIPAALTP-----EFLALGSAGDTV--TVFGLPWINDYSG 231
QY 234 TVLPVLVSWILATIEKFLHKLMTADFLITPVLTLTGLTFLTAIGPANKWGDLLA 293
Db 232 QVFPPLIAAGLYWYEKGLKPIEAVQMVFPFFSLLIMIPATAFLGPGFVGNGIS 291
QY 294 HGLQGLYDFGGPVGGLFGLVYSPVITGLHQSFPPLELEFNQ--GGSFIFATASMANIA 352
Db 292 NLEANNPSPFISIVIPLLYFLVPLGLHFWLNAINQNTLGYDFDIQGMGAWNPA 351
QY 353 QGAACLVAFFLAKSEKLGLAGAS--GVSAYL--GITEPAIFGVNLRWPFYIGITAA 408
Db 352 CFGLVTVGFLLSIKERNKAMRQVSLGMLAGLGGISEPSLXGVLRLFRKFTYRLLPGCL 411
QY 409 IGGALIALFDIKAVAGAGFLGVVSDAPDMVNFVCAVTVFVIFAGAAIAYGLYLVR 467
Db 412 AGIVMGIFDIKAYAVFVLSLLTIPAMD--PWLYTIGIAYAVFVSNFLVLDY----R 465
QY 468 RNSGID-----PDAT-AAAPV-----PAGTKKAEAPAA 494
Db 466 SNEERDEARAKVAADKQAEEDLKAEANATPAAPVAAAGAGAGAGAGAAAGATAAKPK 525
QY 495 EFSNDSTIQAPLTGEAIALSSVSDAMFASGKLGSCVAIVPTKQLVSPVSKIVVAPPS 554
Db 526 LAAGEVVDIVSPLEGRKAIPLEVPDIPFAAGKLGPGIAIOPTGNTVVPADATVILVQKS 585
QY 555 GHAFVAVTRAEQDSNVDILMHIGFTVNLNGTHFNPLKKQGDGKAGELLCEFDIDAIA 614
Db 586 GHAVLRL---DSGVEILLVHVGDLTVQLGGEGFTVHVERQQVKAGDPLITFDADFIRS 641
QY 615 AGYEVTTPVIVSNYKKTGPVNTYGLGEIEAGANLLNVAKK 654
Db 642 KDLPLITPVVYSNAKFEIEGIPADQANSSTTVIKVNGK 681

RESULT 14

```



RA Steinmetz M.;  
 RT "Nucleotide sequence of the sacS locus of *Bacillus subtilis* reveals  
 the presence of two regulatory genes.";  
 RL Gene 90:153-155(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=95020537; PubMed=7934828;  
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,  
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moser I.,  
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,  
 RA Rapoport G., Danchin A.;  
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97  
 kb region from 325 degrees to 333 degrees.";  
 RL Mol. Microbiol. 10:371-384(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codan J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Enrian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghm S.F., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Gusepki G., Guy B.-J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
*subtilis*.";  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP SEQUENCE OF 348-459 FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=92216127; PubMed=1806041;  
 RA Glaser P., Kunst F., Debatbouille M., Vertes A., Danchin A.,  
 RA Dedonder R.;  
 RT "A gene encoding a tyrosine tRNA synthetase is located near *sacS* in  
*Bacillus subtilis*.";  
 RL DNA Seq. 1:251-261(1991).  
 CC -!- FUNCTION: Negatively regulates *sacY*.  
 CC -!- SIMILARITY: Contains 1 PTS EIIB domain.  
 CC -!- SIMILARITY: Contains 1 PTS EIIC domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 CC EMBL; M29333; AAA75335.1; -;  
 CC EMBL; X52480; CAA36719.1; -;

DR EMBL; Z99123; CAB15867.1; -;  
 DR EMBL; X73124; CAA51570.1; -;  
 DR PIR; J00293; J00293.  
 DR HSSP; P05053; 1IBA.  
 DR Subtilist; BG10560; sacX.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR PROSITE; TIGR00826; EIIB\_glc; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Transcription regulation; Transferase; Phosphorylation;  
 KW Complete proteome.  
 FT DOMAIN 1 ? EIIB DOMAIN.  
 FT DOMAIN ? 459 EIIC DOMAIN.  
 FT MOD\_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 136 136 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 433 433 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 459 AA; 49024 MW; A5C4E996EGDAD3D40 CRC64;  
 Query Match 21.1%; Score 706; DB 1; Length 459;  
 Best Local Similarity 35.9%; Pred. NO. 7.9e-36;  
 Matches 168; Conservative 93; Mismatches 167; Indels 40; Gaps 11;  
 QY 3 HKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDKVDVRSQSLDDPDLKGTFFGGMFQ 62  
 DB 2 HREIAKELLLAGGKNIIISICHTRLRDFDKDETKIDHAIENLOGVQGFYRGLFQ 61  
 QY 63 ITVPGDGDVDFKEL-----DDATSKDIAVSTQLKXDVANNWFSRAVKVLADIFVPLI 118  
 DB 62 IIFGAGVKNKIYKVHVHETAPSEE---PVHQKK--ASRKLNPAAAFKLTSDIFVPII 116  
 QY 119 PILVGGGILLMAINNVLVADLFGPQSLVEMFQISGVAEMINLMASAPAFPLVVGFTA 178  
 DB 117 PAITASGLLMG-----LIGMIKVFHWFAAGSPWIKMDLVSTAFILPLVGFSA 167  
 QY 179 TRFRGNEFLGAGIGMAMVFTLVNGDYVAATMTAGEMPMSLFLGLDVAQAGYQGTVLV 238  
 DB 168 ARQFGSNPYLGAIVAGLLTHPDLDP--SMLGSKTSSLDI---GLHIPMGYQGSMP 223  
 QY 239 LVVSVTLATIEKFLHKLKMGTDFTLTPVTLTLTLTGFTFIAIGPAMRWVGDLAHLQ 298  
 DB 224 LUSVFVMSKIEKLLKSIVPKSLDVLPITVITVWVTCGLALIVNPAASIIQIMTQSIYV 283  
 QY 299 LYDFGPGVGGLLFGLVYSPIVITGLHQSPPIELELF---NQGGSFIFATASMANIAQGA 355  
 DB 284 IYDHAGIAAGALFGGIYSTIVLSGLHHSFVAIEATLLANPHVGVNPLVPIWMSMANVAQGG 343  
 QY 356 ACLAVFFLAKSEKLLKLAGASGVSAVLGITEPAIGVNLRLRWPYIGIGTA---AIGA 412  
 DB 344 AGLAVFLKTKQSSLLKIALPASLTAFLGIVFPIGVNKLIRPF---IGAAGGAIGA 400  
 QY 413 LIALFDIKAVCALGAAGFLGVVSDADP-----DMVMFLVCVVTVFV 452  
 DB 401 YVAVQVWANSYGLTG-IPMISIVLPFGAANFVHYMIGELIAAVSAFI 447

Search completed: October 1, 2003, 05:02:27  
 Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 01:21:54 ; Search time 61 Seconds  
(without alignments)  
1719.972 Million cell updates/sec

Title: US-10-019-284A-2  
Perfect score: 3342  
Sequence: 1 MDHKLDAQLRLDIGEDNI.....IEAGNLLNVAKEVPATP 661

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3342	100.0	661	22	AA669080
2	3310	99.0	661	22	AA692650
3	2332	69.8	468	22	AA66707
4	2332	69.8	468	23	ABG80325
5	1806	54.0	362	22	AA665708
6	957	28.6	617	23	ABBA7495
7	956	28.6	627	22	AAU37874
8	949	28.4	655	24	ABP81458
9	943	28.4	655	24	ABU02378

10	935.5	28.0	620	23	ABP27215
11	931	27.9	676	23	ABP25654
12	924.5	27.7	634	23	ABP49833
13	920	27.5	674	23	ABP25655
14	880.5	26.3	639	23	ABP27215
15	859	25.7	617	23	ABBA49923
16	829	24.8	618	23	ABBA7649
17	822.5	24.6	636	23	ABBA4803
18	799.5	23.9	612	24	ABP81336
19	799.5	23.9	612	24	ABU00950
20	791	23.7	622	23	ABP28755
21	785.5	23.5	620	23	ABP26854
22	747.5	22.4	640	23	ABBA49939
23	725.5	21.7	683	22	AAG93207
24	725.5	21.7	683	22	ABG6721
25	725.5	21.7	683	23	ABG80333
26	719	21.5	494	23	ABBA4322
27	691	20.7	704	22	AAU60962
28	584	17.5	381	24	ABP81450
29	558.5	16.7	791	23	ABP66240
30	550	16.5	484	22	AAU35216
31	538	16.1	455	22	AAU38201
32	529	15.8	334	24	ABU02198
33	528.5	15.8	243	23	ABP3269
34	522	15.6	484	22	AAU36679
35	517	15.5	484	24	ABJ19011
36	511	15.3	521	23	ABP40220
37	511	15.3	478	23	ABP53735
38	490.5	14.7	439	22	AAU34164
39	477	14.3	461	23	ABP26853
40	440.5	13.2	454	23	ABP54447
41	438	13.1	474	22	AAU34650
42	422.5	12.6	413	22	ABG66722
43	404	12.1	523	22	AAU46075
44	374	11.2	249	24	ABP81451
45	374	11.2	249	24	ABU02199

ALIGNMENTS

RESULT 1

AA669080  
ID AAB69080 standard; Protein; 661 AA.

XX AAB69080;

XX AC

XX 20-APR-2001 (first entry)

DT Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.

DE Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;

KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;

KW coryneform bacterium; phosphoenolpyruvate-sugar transport system.

XX Brevibacterium lactofermentum.

OS Brevibacterium lactofermentum.

XX WO200102584-A1.

XX 11-JAN-2001.

PD 30-JUN-2000; 2000WO-JP04348.

PF 02-JUL-1999; 99JP-0189512.

PR (AJIN ) AJINOMOTO CO INC.

PA Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;

XX WPI; 2001-138150/14.

DR N-PSDB; AAF32543.

XX Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme

PT

PT II obtained by cassette ligation-mediated amplification of downstream  
 PT domain of coryneform bacterium sucrose gene, with sucrose-binding  
 PT activity -  
 XX  
 PS Claim 1; Page 29-32; 45pp; Japanese.  
 XX  
 CC The present sequence represents the Brevibacterium lactofermentum  
 CC sucrose PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system  
 CC or phosphoenolpyruvate-sugar transport system) enzyme II, which has  
 CC sucrose-binding activity. A coryneform bacteria produced with the  
 CC sucrose-PTS enzyme II gene can have more efficient sugar uptake and  
 CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene  
 CC and its disrupted gene, such as one without the sucrose PTS function,  
 CC can be used to produce new breeds of coryneform bacterial strains to  
 CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can  
 CC have improved amino-acid and nucleic acid productivity.  
 XX  
 SQ Sequence 661 AA;

Query Match 100.0%; Score 3342; DB 22; Length 661;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 661; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MDHKDLAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60  
 DB 1 MDHKDLAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60  
 QY 61 FOIIVGPGDGVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120  
 DB 61 FOIIVGPGDGVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120  
 QY 121 LVGGGLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVVGFTATK 180  
 DB 121 LVGGGLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVVGFTATK 180  
 QY 181 RFGCNFELGAGIGMAVFTFLVNGYDVAATATAGEMPMSLFGDVAQAGYOGTVLPVLY 240  
 DB 181 RFGCNFELGAGIGMAVFTFLVNGYDVAATATAGEMPMSLFGDVAQAGYOGTVLPVLY 240  
 QY 241 VSNILATIEKFLKRLMGTAFLITPVLTLLTGFTTATGPAWRWGDLLAHGLQGLY 300  
 DB 241 VSNILATIEKFLKRLMGTAFLITPVLTLLTGFTTATGPAWRWGDLLAHGLQGLY 300  
 QY 301 DFGPVGGLLFGLVYSPVITGLHQSPFPIELFNQGSFIFATASMANIAQGAACLA 360  
 DB 301 DFGPVGGLLFGLVYSPVITGLHQSPFPIELFNQGSFIFATASMANIAQGAACLA 360  
 QY 361 FFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWPFYIGITGTAAGGALIALFDIK 420  
 DB 361 FFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWPFYIGITGTAAGGALIALFDIK 420  
 QY 421 AVAALGAGFLGVVYSIDAPDMVFLVCAVTVTFVIAFGAAIYAGLYLVRNGSIDPDATAAP 480  
 DB 421 AVAALGAGFLGVVYSIDAPDMVFLVCAVTVTFVIAFGAAIYAGLYLVRNGSIDPDATAAP 480  
 QY 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATIALSSVSDAMFASGKLSGVAIVPTKGL 540  
 DB 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATIALSSVSDAMFASGKLSGVAIVPTKGL 540  
 QY 541 VSPVSGKIVVAFPSGHAFVTRKAEQSGNSVDILMHIGFTVNLNGHFNPLKQSGDEKA 600  
 DB 541 VSPVSGKIVVAFPSGHAFVTRKAEQSGNSVDILMHIGFTVNLNGHFNPLKQSGDEKA 600  
 QY 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPNVYGLGEIBAGANLLNVAKKEAVPAT 660  
 DB 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPNVYGLGEIBAGANLLNVAKKEAVPAT 660  
 QY 661 P 661  
 DB 661 P 661

RESULT 2

AAG92650  
 ID AAG92650 standard; Protein; 661 AA.  
 XX  
 AC AAG92650;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 6404.  
 XX  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH67869.  
 XX  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 661 AA;  
 Query Match 99.0%; Score 3310; DB 22; Length 661;  
 Best Local Similarity 98.9%; Pred. No. 0;  
 Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MDHKDLAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60  
 DB 1 MDHKDLAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60  
 QY 61 FOIIVGPGDGVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120  
 DB 61 FOIIVGPGDGVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120  
 QY 121 LVGGGLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVVGFTATK 180  
 DB 121 LVGGGLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVVGFTATK 180  
 QY 181 RFGCNFELGAGIGMAVFTFLVNGYDVAATATAGEMPMSLFGDVAQAGYOGTVLPVLY 240  
 DB 181 RFGCNFELGAGIGMAVFTFLVNGYDVAATATAGEMPMSLFGDVAQAGYOGTVLPVLY 240

Db 181 RFGNEFLCAGIGMVFPSLVNGYDVAATMAAGEMPWMSLFGLDVAAGYQGTVPVPLV 240  
 Qy 241 VSNLATIEKFLHKLMTADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLQGLY 300  
 Db 241 VSNLATIEKFLHKLMTADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLQGLY 300  
 Qy 301 DFGPGVGLLGLVYSPVITGLHQSFPPPIELFNOGGSFIFATASMANIAQGAACLA 360  
 Db 301 DFGPGVGLLGLVYSPVITGLHQSFPPPIELFNOGGSFIFATASMANIAQGAACLA 360  
 Qy 361 FFLAKSEKLGAGAGSVSAVLGITPAIFGVNLRWRPPIGIGTAAGGALLIALFDIK 420  
 Db 361 FFLAKSEKLGAGAGSVSAVLGITPAIFGVNLRWRPPIGIGTAAGGALLIALFNIX 420  
 Qy 421 AVALGAAGFLGVYSDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRNGSIDPDATAAP 480  
 Db 421 AVALGAAGFLGVYSDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRNGSIDPDATAAP 480  
 Qy 481 VPAGTTKAEAEAPERSNDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGOL 540  
 Db 481 VPAGTTKAEAEAPERSNDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGOL 540  
 Qy 541 VSPVSGKIYVAFPSGHAFVTRKAEDEGSNDIILMHIGFDTVNLNGTHFNPLKKGDEVKA 600  
 Db 541 VSPVSGKIYVAFPSGHAFVTRKAEDEGSNDIILMHIGFDTVNLNGTHFNPLKKGDEVKA 600  
 Qy 601 GELLCEFDIDAIAKAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 660  
 Db 601 GELLCEFDIDAIAKAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 660  
 Qy 661 P 661  
 Db 661 P 661

## RESULT 3

AA66707  
 ID AAB66707 standard; protein: 468 AA.

XX AC AAB66707;

XX DT 09-APR-2001 (first entry)

XX DE C-glutamicum phosphoenolpyruvate protein #1.

XX KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX OS Corynebacterium glutamicum.

XX PN W0200102583-A2.

XX PD 11-JAN-2001.

XX PF 27-JUN-2000; 2000WO-IB00973.

XX PR 01-JUL-1999; 99US-0142691.

XX PR 23-AUG-1999; 99US-0150310.

XX PR 03-SEP-1999; 99DE-1042095.

XX PR 03-SEP-1999; 99DE-1042097.

XX PA (BADI ) BASF AG.

XX PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

XX XX WPI; 2001-080989/09.

XX XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;  
 PT sugar phosphotransferase system proteins or their portions, useful for  
 PT typing or identifying C. glutamicum or related bacteria, and as markers  
 PT for transformation -

XX PS Claim 4; Page 101-102; 144pp; English.

XX PS

CC The present invention relates to Corynebacterium glutamicum  
 CC phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.  
 CC The PTS nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or  
 CC related bacteria, the typing or identification of C. glutamicum or  
 CC related bacteria, as reference points for mapping C. glutamicum genome,  
 CC and as markers for transformation.

XX Sequence 468 AA;

Qy Query Match 69.8%; Score 2332; DB 22; Length 468;  
 Best Local Similarity 98.5%; Pred. No. 7.2e-223;  
 Matches 461; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 194 MAWVPTLVNGYDVAATMAAGEMPWMSLFGLDVAAGYQGTVPVPLVVSILATIEKFLH 253  
 Db 1 MAWVPSLVNGYDVAATMAAGEMPWMSLFGLDVAAGYQGTVPVPLVVSILATIEKFLH 60  
 Qy 254 KRLMTADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLQGLYDFGPGVGLLFLGL 313  
 Db 61 KRLMTADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLQGLYDFGPGVGLLFLGL 120  
 Qy 314 VYSPVITGLHQSFPPPIELFNOGGSFIFATASMANIAQGAACLAFFLAKSEKLGKLA 373  
 Db 121 VYSPVITGLHQSFPPPIELFNOGGSFIFATASMANIAQGAACLAFFLAKSEKLGKLA 180  
 Qy 374 GASGVSAVLGITPAIFGVNLRWRPPIGIGTAAGGALLIALFDIKAVAGAGFLGVV 433  
 Db 181 GASGVSAVLGITPAIFGVNLRWRPPIGIGTAAGGALLIALFDIKAVAGAGFLGVV 240  
 Qy 434 SIDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRNGSIDPDATAAAPPAGTTKAEAP 493  
 Db 241 SIDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRNGSIDPDATAAAPPAGTTKAEAP 300  
 Qy 494 AEFSDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKITVAVFP 553  
 Db 301 AEFSDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKITVAVFP 360  
 Qy 554 SGHAFVTRKAEDEGSNDIILMHIGFDTVNLNGTHFNPLKKGDEVKAGELICEFDIDA 613  
 Db 361 SGHAFVTRKAEDEGSNDIILMHIGFDTVNLNGTHFNPLKKGDEVKAGELICEFDIDA 420  
 Qy 614 AAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 661  
 Db 421 AAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 468

## RESULT 4

ABG80325  
 ID ABG80325 standard; Protein: 468 AA.

XX AC ABG80325;

XX DT 15-NOV-2002 (first entry)

XX DE C. glutamicum metabolic pathway (MP) protein #5.

XX KW Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;  
 KW cofactor; nucleotide; nucleoside; trehalose; fine chemical production;  
 KW organic acid; non-proteinogenic amino acid; purine base; carbohydrate;  
 KW pyrimidine base; lipid; unsaturated fatty acid; diol; polyketide;  
 KW aromatic compound; food industry; animal feed; cosmetic industry;  
 KW pharmaceutical industry; enzyme.

XX OS Corynebacterium glutamicum ATCC 13032.

XX PN W0200251231-A1.

XX XX 04-JUL-2002.

XX PS 22-DEC-2000; 2000WO-EP13143.

XX PS

```

PR 22-DEC-2000; 2000WO-EP13143.
XX (BADI ) BASF AG.
PA Pompejus M, Kroeger B, Zelder O, Schroeder H;
PI WPI; 2002-643289/69.
XX N-PSDB; ABS65346.
XX New metabolic pathway genes of Corynebacterium glutamicum for producing
PT fine chemicals, e.g. lipids, (unsaturated fatty acids, vitamins,
PT cofactors or enzymes used in food, feed, cosmetics or pharmaceutical
PT industries -
XX Claim 18; Page 111-113; 176pp; English.
XX The present invention relates to the isolation of Corynebacterium
CC glutamicum metabolic pathway (MP) proteins, and the polynucleotide
CC sequences encoding them. The MP proteins are enzymes involved in
CC the metabolism of molecules important for the normal functioning
CC of cells (e.g. amino acids, vitamins, cofactors, nucleotides and
CC nucleosides, or trehalose). The polynucleotide sequences encoding
CC the MP proteins are useful for producing fine chemicals, particularly
CC organic acids, non-proteinogenic amino acids, purine and pyrimidine
CC bases, nucleosides, nucleotides, lipids, (unsaturated fatty acids,
CC diols, carbohydrates, aromatic compounds, vitamins, cofactors,
CC polyketides and enzymes). The fine chemicals are useful in the food,
CC animal feed, cosmetic or pharmaceutical industries. ABG80321-ABG80343
CC represent the C. glutamicum MP proteins of the invention.
XX Sequence 468 AA;
SQ
Query Match 69.8%; Score 2332; DB 23; Length 468;
Best Local Similarity 98.5%; Pred. No. 7.2e-233;
Matches 461; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
QY 194 MANVFTLVNGYDVATMTAGEMPMSLFGDLVAQAGYGTGVLPLVWSWILATIEKFLH 253
DB 1 MANVFTLVNGYDVATMTAGEMPMSLFGDLVAQAGYGTGVLPLVWSWILATIEKFLH 60
QY 254 KRLMGADFLITPVLPLLTGTLTFTTAIGPAMRWGDLAHLQGLYDFGGVGLLFLG 313
DB 61 KRLMGADFLITPVLPLLTGTLTFTTAIGPAMRWGDLAHLQGLYDFGGVGLLFLG 120
QY 314 VYSPVITGLHQSPPIELFNOGGSFIFATASMANIAQGAACIAVFFLAKSEKLGIA 373
DB 121 VYSPVITGLHQSPPIELFNOGGSFIFATASMANIAQGAACIAVFFLAKSEKLGIA 180
QY 374 GASGVSAILGITEPAIFGVNLRWPFYIGIGTAAIGGALIALFDIKAVAGLGVV 433
DB 181 GASGVSAILGITEPAIFGVNLRWPFYIGIGTAAIGGALIALFNKAVAGLGVV 240
QY 434 SIDAPDMVFLVCVVTFIATGAATAYGLYLVRNGSIDDPATAPVPAGTTKAEAP 493
DB 241 SIDAPDMVFLVCVVTFIATGAATAYGLYLVRNGSIDDPATAPVPAGTTKAEAP 300
QY 494 AEFNSDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKGOLVSPVSGKIIVAFP 553
DB 301 AEFNSDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKGOLVSPVSGKIIVAFP 360
QY 554 SGHAFVRKAEKGSNDILMHIGFTVNLNTHFNPLKKQDEYKAGELICEFDIDA 613
DB 361 SGHAFVRKAEKGSNDILMHIGFTVNLNTHFNPLKKQDEYKAGELICEFDIDA 420
QY 614 AAGYEYVTTPIVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 661
DB 421 AAGYEYVTTPIVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPATP 468
RESULT 5
AAB66708
ID AAB66708 standard; protein; 362 AA.
XX

```

```

AC AAB66708;
XX 09-APR-2001 (first entry)
XX C.glutamicum phosphoenolpyruvate protein #2.
XX Phosphoenolpyruvate; sugar phosphotransferase system; Pts.
XX Corynebacterium glutamicum.
XX WO200102583-A2.
XX 11-JAN-2001.
XX 27-JUN-2000; 2000WO-IB00973.
XX 01-JUL-1999; 99US-0142691.
XX 23-SEP-1999; 99US-0150310.
XX 03-SEP-1999; 99DE-1042095.
XX 03-SEP-1999; 99DE-1042097.
XX (BADI ) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-080989/09.
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -
XX Claim 6; Page 104-106; 144pp; English.
XX The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins which are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.
XX Sequence 362 AA;
Query Match 54.0%; Score 1806; DB 22; Length 362;
Best Local Similarity 99.2%; Pred. No. 1.1e-170;
Matches 359; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 300 YDFGPGVGLLGLVYSPVITGLHQSPPIELFNOGGSFIFATASMANIAQGAACIA 359
DB 1 YDFGPGVGLLGLVYSPVITGLHQSPPIELFNOGGSFIFATASMANIAQGAACIA 60
QY 360 VFFLAKSEKLGAGASVSALGITTEPAIFGVNLRWPFYIGIGTAAIGGALIALFDI 419
DB 61 VFFLAKSEKLGAGASVSALGITTEPAIFGVNLRWPFYIGIGTAAIGGALIALFNI 120
QY 420 KAVAGLGAAGFLGVYSIDAPDMVFLVCVVTFIATGAATAYGLYLVRNGSIDDPATA 479
DB 121 KAVAGLGAAGFLGVYSIDAPDMVFLVCVVTFIATGAATAYGLYLVRNGSIDDPATA 180
QY 480 PVPAGTTKAEAPAEFNSDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKQ 539
DB 181 PVPAGTTKAEAPAEFNSDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKQ 240
QY 540 LVSPVSGKIIVAFPSSGHAFVRKAEKGSNDILMHIGFTVNLNTHFNPLKKQDEYK 599
DB 241 LVSPVSGKIIVAFPSSGHAFVRKAEKGSNDILMHIGFTVNLNTHFNPLKKQDEYK 300
QY 600 AGEELICEFDIDA KAGYEYVTTPIVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 659
DB 301 AGEELICEFDIDA KAGYEYVTTPIVSNYKKTGPVNTYGLGEIEAGANLLNVAKKEAVPA 360

```

QY -660 TP 661  
 II  
 Db 361 TP 362

RESULT 6  
 ABB47495  
 ID ABB47495 standard; Protein; 617 AA.  
 XX  
 AC ABB47495;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #199.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 FN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain I, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia del Portillo F, Gomez-Lopez N;  
 PI Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 DR WPI; 2002-010914/01.  
 XX  
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -  
 XX  
 PS Claim 6; SEQ ID No 200; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present invention is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 617 AA;  
 Query Match 28.6%; Score 957; DB 23; Length 617;  
 Best Local Similarity 35.1%; Pred. No. 8.4e-86;  
 Matches 225; Conservative 124; Mismatches 226; Indels 64; Gaps 17;  
 QY 1 MDHKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDKVDQRSLDDDDPDLKGFETGGM 60

Db 1 MDYQKLAEILANVGGEENRVSVHCATRLRFKLVNKEKADKKQIESISGYSIVVENAQG 60  
 QY 61 FQIIWG--PGDVHDV---FKELDDATSKDIAVSTEQKLDVVANNANFSAVKVLADIFV 115  
 Db 61 LQVILGNVGDVYKALGSFTKLTDDGSEIAKGT---KD---SGNFSLKRAIDVISGIFT 114  
 QY 116 PLIPLVGGGLMAINNVYVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAELPLVVG 175  
 Db 115 PILGALAGGMLKGLLMILT---FG-----WLTESGTYQILYAADSIFYELPLILA 165  
 QY 176 FTATKRFEGNEFLGAGICGMAMVFTLVNGYDVAATMTAGEMPMSLFLGLDVAQAGYQGV 235  
 Db 166 YTAARKFGANPPVATAAGALVYPTMINLFNEGAHIFLQIP-----VILMSYSFSV 217  
 QY 236 LPVLVYSILATIEKFLHRLMGTAFLITPVLTLTLTGLTFTFAIGPAMRWGDLAHLG 295  
 Db 218 IPILAVWFSLERFLNSKIHEAAKTELTPEMCLMLIVPLTLFAFGPLGTFISOGLASG 277  
 QY 296 LQGLYDFGPGVGLLFLGLVYSPITVITGLHQSPPIEL-ELFNQGSFFIFATASMANIAQG 354  
 Db 278 YTFIYNLSPIVAGAFMGAFQVQLVIFGIHWGFPVIMINNSRYGDRDTMIAMVGPNSFAQA 337  
 QY 355 AACLAVFFLAKSEKLGKLAGASGVSALVIGTEPAIFGVNLRWRPFYIGICTAAIGGALI 414  
 Db 338 GASLGVFLKTKPEVKAIAGSAALTGFFGITEPSIYGVTLKYKKDFVIASITAGATGAIV 397  
 QY 415 ALFDIKAVALGAAGFLGVVSDIDAPDV---MFLVCVVTFVIAFGAAIAY-----GLYIV 466  
 Db 398 -----GAAGSSGAANA-IPGILTLPIFGKGVGFIL--GIAYAILLSALTGTF 444  
 QY 467 RRGNSIDPDATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASCK 526  
 Db 445 ---GYKDEKADGI---APTKEAKETGVE---AEVIVPIRGNTVPLNEVKDEAFSAGL 494  
 QY 527 LSGVAIVPTKQLVSPVSGKIVAFPSGHAPAVETKAEKDSNVDLMHIGEDTYNLNGT 586  
 Db 495 LCKGVAIVPQEGKLISPVNGTITETAPPTGHAIGIRS----DKGVEILLHVGEDTVOLNKG 550  
 QY 587 HFNPLKKQGDVEKAGELLCEFDIDAIAKAAGYEVTTPIVVSN 627  
 Db 551 YFKLLVAQGDRLVVGQALLFQLEAKADGYDITPIVVTN 591

RESULT 7  
 AAU37874  
 ID AAU37874 standard; Protein; 627 AA.  
 XX  
 AC AAU37874;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Streptococcus pneumoniae cellular proliferation protein #303.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-20727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX

[illegible]



QY 357 CLAVFLAK-SEKLGKLAGASGVSAVIGITEPAIFGVNLRWPYIIGTAAIGALIA 415  
 Db 362 VFAYFMHRRHREQAQVSLPATISAYLGVTEPAIFGVNKKYIYPVAGMTGSALAGLSV 421  
 QY 416 LFDIKAVALLGAAGFLGVVSIDAPDMVF-----LVCAVTVFVIAFGAAIAYGLYLVRNGS 471  
 Db 422 TENVTAAASIGIGLPGILSIQPYMLPFGAGTMLVAIVPMLTF-----PPRKAGL 472  
 QY 472 IPDPTAAAPVPAQTAKAE-----APAEFSDSTI-----IQAPLTGEATLSVSDAM 521  
 Db 473 FTK-----TEGDTNLOAEFVAQEAEEFVNEPVELTSVELISPLTGQVKELSQATDPI 525  
 QY 522 FASGKLGSGVAIVPTKQGLVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDILMHIGFDIV 581  
 Db 526 FASGVMGQGLVIERPSOGELTSPVNGTIVFLPPTKHAIGI--VSDG--VELLIHIGMDIV 581  
 QY 582 NLNGHFNLKKGDEVKAGELLCBFDIDAIAKAAGYEVTTPIVNSN---YKKTGPVNTYG 638  
 Db 582 GLDGKGFESLVVQGDHVTYQQLIRPDMDIVKAAGLVTTETPVITNQDAYTATIP-GTYP 640  
 QY 639 LGEIEAGANLL 649  
 Db 641 T-TIAGASIM 650

## RESULT 10

ABP27216  
 ID ABP27216 standard; Protein; 620 AA.

XX AC ABP27216;  
 XX DT 02-JUL-2002 (first entry)  
 DE Streptococcus polypeptide SEQ ID NO 3508.  
 XX KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX OS Streptococcus pyogenes.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WP1; 2002-352536/38.

XX DR N-PSDB; ABN67847.

XX PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -

XX PS Claim 1; Page 3514; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

XX SQ Sequence 620 AA;

Query Match 28.0%; Score 935.5; DB 23; Length 620;  
 Best Local Similarity 34.7%; Pred. No. 1.2e-83;  
 Matches 226; Conservative 108; Mismatches 273; Indels 45; Gaps 11;

QY 1 MDHDLAQRILRDITGGEDNIVAAAHCAATRLRLVLKDTKDVDRSLDDDDPOLKTFETGGM 60  
 Db 1 MNRQIAAEVTEALGGRENRSVAHCAATRLRLVMDYDEKIDKEKAAIDKVKGAFFNSGQ 60  
 QY 61 FOIIVGPDGDVDFVKELDDATSKDIIV-----STEOLKDVVANNANWFSRAVKVLADIFV 115  
 Db 61 YQMIFGTGTYNNIYDEV-----VALGLPTSSTSEKAEAGKGNIFQRAIRTFGDFV 113  
 QY 116 PLIPILVGGGLLMAINNVILVAQ-----DLFGPQSLVEMFPQISGVAEMINLMASAPFALP 171  
 Db 114 PIIPAIIVATGLFMGVRG-LVTQPAIMDLFGVHEYGENFLMYT-----RLTDTAFVYLP 166  
 QY 172 VLVGFTATKRFNGEFLGAGIGMAWFTLVNGVDVAATMTAGEMWMSLFLGLDVAQAGY 231  
 Db 167 ALVWASAFRVGGNPIGIVGLMLVSNELPNWVVA--SGGDVKLTFFGF-VPVVG 222  
 QY 232 QGTVLPLVLSMILATIEKFLHKLMTADFLTPVLTLLLTGLTFLTIALGPMRWGDL 291  
 Db 223 QGTVLPAFFVLGAKLEKHLKVPKPEALDLVTPFTFAIMSTLGLFVIGPVFHSLENL 282  
 QY 292 LAHGLQGLYDFGPGVGLLGLVSPVITGLHQSPFPIELELF-NOGGSFIFATASMAN 350  
 Db 283 VLAGTQAVLHLPFGIAGLIVGGIQLIVVTGIHIFNLEAQLIANTGKDPFNAYLTAAT 342  
 QY 351 IAQGAACLVAFFLAKSEKLGKLAGASGVSAVIGITEPAIFGVNLRWPYIIGTAAIG 410  
 Db 343 AAQAGATLAVAVKTKSTKLGLAPFPTSALLGLITEPAIFGVNLRPKVFSVGLIGGALG 402  
 QY 411 GALLALFDIKAVALGAAGFLGVSIDAPDMVMFLVCAVTVFVIAFGAAIAYGLYLVRNG 470  
 Db 403 GWVAGLFGIAGTGFGIIVLPOTLLYLNQQLLYLTMVLGLVAFATAYTNGY----- 455  
 QY 471 SIDPDATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEAIALSVSDAMFASGLKLGSG 530  
 Db 456 ---QDRETPLPFAVEVDQTADQALAE---TLXSPLNGTVVDLSAVSDPVFSSGANGOG 509  
 QY 531 VAIPTKGQLVSPVSGKIIVAFPSGHAFVARTKAEDGSNVDILMHIGFDIVNLNGTFNP 590  
 Db 510 LAIPEDNTLSPVDGKRVETFTGHAYTIS-----SQGAELVLIHIGIDTESMAGDFES 565  
 QY 591 LKKGDEVKAGELICEFDIDAIAKAAGYEVTTPIVVSNNYKKTGPVNTYGLGEI 642  
 Db 566 LVAVGQAVKKGDLGLGHFDPKIAEAGLDLDTMTMIVSNIAIYQSVDIILAQOHV 617

## RESULT 11

ABP25654

ID ABP25654 standard; Protein; 676 AA.

XX AC ABP25654;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 484.

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX OS Streptococcus agalactiae.  
 XX PN W0200234771-A2.  
 XX PD 02-MAY-2002.  
 XX XX 29-OCT-2001; 2001WO-GB04789.  
 XX PF 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.  
 XX XX (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
 XX PI Tettelin H;  
 XX XX WPI; 2002-352536/38.  
 XX DR N-PSDB; ABN66285.

XX PT New Streptococcus protein for the treatment or prevention of infection  
 XX or disease caused by Streptococcus bacteria, such as meningitis, and  
 XX for detecting a compound that binds to the protein -  
 XX PS Claim 1; Page 3202-3203; 4525pp; English.  
 XX XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX SQ Sequence 676 AA;

Query Match 27.9%; Score 931; DB: 23; Length 676;  
 Best Local Similarity 34.6%; Pred. No. 3.8e-83;  
 Matches 237; Conservative 117; Mismatches 277; Indels 54; Gaps 17;  
 QY 7 AQRILRDIGGDNIVAAHCAATRLRLVLDKVDVDSQSLDDPLDKTFTGGMFOIIVG 66  
 DB 8 AKALLEAIGGKENSATVTHCATRFRVLDNSKAKVIEELPSVKGTFTNAGQFVIIG 67  
 QY 67 PGDVDHVFKELEDDATSKDIAVSTEQLDKDWANNWFSAVKVLADIFVPLIPLVGGGL 126  
 DB 68 -NDVP-IPYNAFVAVSGIEGYSKAAKQKQNLQRLVLTMLAEITPIPIALIVGGL 125  
 QY 127 LMAINNVLAQDL-FGQSLVEMFPQI-----SGVAEMINLMASAPF 167  
 DB 126 ILGFENILDAVFFELGQKVVGVQVDSSGHPITWTLVDVSTWGSVDFLWLPGBAIF 185  
 QY 168 AFLPVLVFTATKFRGGNEFLGAGIGAMVFPFTLVNGYDVAAATAGEMPWSL-FG-LD 225  
 DB 186 HFLPVGIVNSYTRKMTTQILGIVLGVLPQLLNAYSVASTSAAADIAKNSWNFCYFT 245  
 QY 226 VAQGYQGTVLPLVNVSHLTIETKFLHKLMTGADFLITPVLTLTGTFTFAIGP-- 293

DB 246 VOKICYQAOVIPALLAGLSLSYLEIFWFKKHPEVVSMIFVFLSLVPAIIIAHTVLGPIG 305  
 QY 284 --AMRWGDLIAHGQGLYDFGGPVG---GLLFLGVYSPVITVGHQSPFIELELFNQG 338  
 DB 306 WTLGKWISAIVLIGLT-----GPVKWLFGAIFGALYAPFVITGLHHTNAIDTQLIADI 359  
 QY 339 GSF--IFATASMANIAOGAACLAVFFLAK-SEKLKLAGAGSGVSAVIGITEPALFVGNL 394  
 DB 360 KTHTTGLWPMIALSNAOGSAVLAIYFMRHDEKEAQISLPAAISAYLGVTEPALFGVNV 419  
 QY 395 RLRWPFYIGTAAIGGALIALFDIKAVAGAGFLGVVSDIDAPDMVMFLYCAVVFVIA 454  
 DB 420 KVIYFVAGMIGSSVAGLLATTFNVQANSIGVGGLPGLFSINVKVMGYFFICMAVAIIP 479  
 QY 455 FGAATAY---GLYIVRNGSIDPDATAAPVPAGITKAE-AEAPAFSNDSTIIQAPLTGE 510  
 DB 480 LFLTLLFFKSGILTRTEEEKLVDA-----VIASITETSAKEKAVVSTKLSVSPSLGL 535  
 QY 511 AIALSVSDAMFASGLASGVAIVPTKGQLVSPVSGKIVVAFPPSGHAFVARTKABDGSNV 570  
 DB 536 AKPLDOASDPVFSQIMKGVVIDPDSGELYSVDATVSVLFPTRKHAIGLLT----SEGV 591  
 QY 571 DILMHIGFTVNLNGTHFNPLKKQGVKAGELLCFEDIDAIAKAGYEVTTPIVVSNNYKK 630  
 DB 592 EFLIHIGMTVNLGEGFTSHVAQGDVTKVSGDKLITFDIPMIKEGYIVETPILITNOQE 651  
 QY 631 TGPVNTYGL-GEIEAGANLLNVAKK 654  
 DB 652 FRPEELIDLPKOIKRQALM-VAKK 675

## RESULT 12

ABB49833  
 ID ABB49833 standard; Protein: 634 AA.  
 AC ABB49833;  
 XX 05-FEB-2002 (first entry)  
 DE Listeria monocytogenes protein #2537.  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 OS Listeria monocytogenes.  
 XX WO200177335-A2.  
 XX PD 18-OCT-2001.  
 PF 11-APR-2001; 2001WO-FR01118.  
 PR 11-APR-2000; 2000FR-0004629.  
 XX (INSP ) INST PASTEUR.  
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durand L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides  
 XX PS Claim 6; SEQ ID NO 2538; 192pp; French.  
 XX XX

CC The present invention relates to the genome sequence of *Listeria*  
 CC monocytes EGB-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC monocytes and related organisms, and for studying genetic  
 CC polymorphisms and related genomes. The present invention is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies. Identification of *L.* monocytes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L.* monocytes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by *L.*  
 CC monocytes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 634 AA;  
 Query Match 27.7%; Score 924.5; DB 23; Length 634;  
 Best Local Similarity 32.7%; Pred. No. 1.5e-82;  
 Matches 217; Conservative 126; Mismatches 229; Indels 91; Gaps 16;  
 QY 1 MDHKLDAORILRDIGEDNIVAAHCAATRLRLVLDKTDVDRQSLDDPDLKGTFTGGM 60  
 DB 1 MKYEQLADILKNVGGKENSINVFHCITSLRFLKLDENIANTKETEKLGDVVISIKSGG 60  
 QY 61 FOIIVPGDGDHVEF-----ELDDASKDIAVSTEQLDKDVANNANWFSRAVKL 110  
 DB 61 YQVVG-NHVPDVFRAVLGVGISAEGDGS--APAT-----GNIFRFIDMI 106  
 QY 111 ADIFVPLIPILVGGILLMAINVLVAQDLFGPQSILVEMFPQISGVAEMINLMASAPPAFL 170  
 DB 107 SGVFTPVGLAATGMKGFAMFA--FG-----WITVTSQTYQLLYAIGDCLFYFF 157  
 QY 171 PVLVGTAKRGGNEFLGAGTGMWVPTLVNGVDVAATMAGMPWMSLF----- 222  
 DB 158 PFLGTAKKRGGNIFGMAIGGALVPTL-----AGTAGD-PLYTLFAGTIFRSP 209  
 QY 223 -----GLDVAQAGYQGTVPVLVYVSWILATIEKFLHKLMTADFLITPVLTLTGL 276  
 DB 210 IHVTFGLGIPVILMSYASSIPIVATYFGSKVEGFKKIIPDVIKTFVVPFCTLLIVPI 269  
 QY 277 TETATGPMRWGDLAHGLQGLYDGGPVGGLLEGLVSPVITGLHQSPPIEL-ELF 335  
 DB 270 TFIIVGPIATWAGQLLAGTIWYNLSPIIAGLIGGFQVVFGLHGLVFPVAINLT 329  
 QY 336 NOGGSFIFATAGMANIAQGAACLAFFFLAKSEKRLGLAGASGSAVLGITEPAIFGVNLR 395  
 DB 330 VLGHDPILAMTFGASPAQAGAVLAVFFKSRKKIKSLSPATISGIFGVTPEAIVGVTLP 389  
 QY 396 LRWPFVIGTGAIGGALLTALDIKAVAGLGAAGFLGVSIDAP-----DMYFLVCAY 449  
 DB 390 LKKPFIMSCIAAGIGGIIIGFAGSQTYIMGGIGIFGLPNFFPKPGSGISGEFVWVIAVI 449  
 QY 450 TFVIAFGAAIAYGL-----YLVRNGSIDPDATAAPVPAGTAKAEAPAEFSNDSTIIQ 504  
 DB 450 SFLGFLIVYVGVKPADVWVEQSNTVE-----GETLIERET-----IP 489  
 QY 505 APTGGAIALSSVSDAMFASGKLGSGVAIVPKGQLVSPVSKIVVAPPSGHAFAVRKA 564  
 DB 490 APVVGIVTLADYKDEAFSSGALGKGAIVPTVGRVYVAPAGTIVTTFPTGHAIGTTRK 548  
 QY 565 EDGSNDVILMHIGHFDVNLNGHFNFLKKGQDVKAGELLCEFDIDAKAAGYEVNTPIV 624  
 DB 549 -DGA--EVLIHIGNDVIQLEKGFETHAVKQGDVIEKQGLLTEDIEGIRAGDYDTPPV 605  
 QY 625 VSN 627  
 DB 606 VTN 608

## RESULT 13

ABP25655  
 ID ABP25655 standard; Protein; 674 AA.  
 XX AC ABP25655;  
 XX DT 02-JUL-2002 (first entry)  
 XX STREPTOCOCCUS POLYPEPTIDE SEQ ID NO 486.  
 XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus pyogenes.  
 OS Streptococcus pyogenes.  
 XX WO200234771-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX WPI; 2002-352535/38.  
 DR N-PSDB; ABN66286.  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX Claim 1; Page 3203; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71536 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX Sequence 674 AA;  
 SQ

Query Match 27.5%; Score 920; DB 23; Length 674;  
 Best Local Similarity 33.6%; Pred. No. 4.7e-82;  
 Matches 222; Conservative 122; Mismatches 256; Indels 60; Gaps 15;  
 QY 7 AQRILRDIGEDNIVAAHCAATRLRLVLDKTDVDRQSLDDPDLKGTFTGGMFQIIVG 66  
 DB 8 AKSLTALGGENKIVVTHCATRMRFLVNDNNKANVKEIKISVVKGTFTNAGQFQIIVG 67  
 QY 67 PGDVDHVEKELDDATSKDIATVSTEQLDKDVANNANWFSRAVKVLAIVPLIPILVGGGL 126  
 DB 68 -NDVP-VFYNDFTAVSSIEGVSKEAKSNQALQRVMTMLAEITPTIPIIVGGGL 125



RESULT 15

ABB49923  
ID ABB49923 standard; Protein; 617 AA.

XX  
XX  
AC ABB49923;

XX  
XX  
DT 05-FEB-2002 (first entry)

XX  
XX  
DE Listeria monocytogenes protein #2627.

XX  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.

XX  
OS Listeria monocytogenes.

XX  
PN WO200177335-A2.

XX  
PD 18-OCT-2001.

XX  
PF 11-APR-2001; 2001WO-FR01118.

XX  
PR 11-APR-2000; 2000FR-0004629.

XX  
PA (INSP ) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusnick C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;

XX  
DR WPI; 2002-010914/01.

XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides

XX  
PS Claim 6; SEQ ID NO 2628; 192pp; French.

XX  
CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 617 AA;

Query Match 25.7%; Score 859; DB 23; Length 617;  
Best Local Similarity 30.3%; Pred. No. 4.8e-76;  
Matches 202; Conservative 140; Mismatches 261; Indels 64; Gaps 14;

QY 1 MDHKLQAQRLRIGGEDNVAACATRLRLVLRKDKVDRSLDDPDLKGTFTGCM 60  
DB 1 MDYNQLAKEILQAVGGKNNVNEVHCITRLRFOLKQSKVNEKQLKGLDKVGTNVAGNO 60

QY 61 FQIIIVPGVDHVPKELDDAT-----SKDIAYSTEQLKDVVANNANWFSRAVKVLADIFVP 116  
DB 61 FQVIIG-NDVPKVFDALEAENPAWKNKTKTKTQKGI-----KGFSEIFSLSGVFAP 115  
QY 117 LIPILVGGGLLMAINNVVAQDLFGQSLVEMEPQISGVA-----EMINLKASAPFAEL 170  
DB 116 ILPAIAGAGLI-----KGFMAFVSLGLWLATDTETRYILLIAGDGVFVEL 160  
QY 171 PVLVGTATKRRFGNBEFLGAGIGAMVFPPLVNGYOVAAATMTAGEMPMSLSFGLDVAQAG 230  
DB 161 PILVAVSAARYFRANMFVALGIGALLYP-----DLTALLSAGTTP--HFICLPVTPVT 212  
QY 231 YOGTVLPELVVSWILATIEKFLHKLMTADEFLITPVLTLTLGLTFTFAIGPAMRWVGD 290  
DB 213 YAYSVIPILLAIWMSYVEKWDRIPTSLKLLFVPLITMFIVVPLTLVVGPLGTFVGD 272  
QY 291 LLAHGLQGLYDFGPGVGLLFGLVYSPITVITGLHQSPPIEL-ELFNQGGSFIFATASMA 349  
DB 273 GVSIGINLLNNGAIGGILIGGAMAIIVMTGHHYAIIVPEVISNLAKYGYDKFLPITYIS 332  
QY 350 NIAOGAACLAFFELAKSEKLGAGASGSAVLGITPEPAIFGVNLRNRWPFYIGIGTAAI 409  
DB 333 NMSQAGATFGVFFRAKDKKLSLAFSGLTALMGVTEPAMYGINVYKRFPMASLIGGAA 392  
QY 410 GGALIALFDIKAVAL-GAAGFLGVVSDADPMVNFVCAVVFVFAFGAAIAIYGLVLR 458  
DB 393 GGGFAMFVGKAYVLTGNGGIPGLVGDTFVYALIAMALAFIALIFSIFIGI-----D 448  
QY 469 NGSIDPDATAAPVPAGTTKAEAPAFSNDSTIIQAPLAGEATLSSVSDAMFASGKLG 528  
DB 449 EQMVEATPVADKIRAGT-----ELIQAPVTIGELVKMSQVNDTTFADIEIMG 493  
QY 529 SGVAIVPTKGOLVSPVSGKIVVAPPSGHAFAVRITKAEDSGSNVDILMHIGFDTVNLNTHF 588  
DB 494 KSIKPKNEGKLYAPANGTIISLEKTKHAIAM--KSDNGA--EILLHVGIDTVKLDGNYF 549  
QY 589 NPLKKQGDVKAGELLCEFDIDAIAAGYEVVTTPIVVSNNYKKTGPVNTYIGIETEAGANL 648  
DB 550 TAHVATGVDVBEQGLLVTFDMEKI-AEKYDTTMMVITNTNEYAVVEAKENGIVTKGNQV 608  
QY 649 LNVAKKE 655  
DB 609 MELRSEQ 615

Search completed: October 1, 2003, 05:01:40  
Job time : 63 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 04:31:10 ; Search time 68 Seconds  
(without alignments)  
2508.423 Million cell updates/sec

Title: US-10-019-284A-2

Perfect score: 3342  
Sequence: 1 MDHKLDAQRILRDIGEDNI.....IEAGANLLNVAKEAVPATP 661

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3310	99.0	661	16 Q8NMD6	Q8nmd6 corynebacte
2	1508	45.1	650	2 Q9S6S6	Q9s6s6 lactococcus
3	1264	37.8	534	2 Q93ML1	Q93ml1 lactobacill
4	1186	35.5	481	16 Q8CN82	Q8cn82 staphylococ
5	1183	35.4	480	16 Q8NV35	Q8nv35 staphylococ
6	1178	35.2	480	16 Q9RQ00	Q9rq00 staphylococ
7	1098.5	32.9	627	16 Q9L8G6	Q9l8g6 clostridium
8	1093	32.7	479	16 Q9KLT8	Q9klt8 vibrio chol
9	1013.5	30.3	470	16 Q9KAS1	Q9kas1 bacillus ha
10	979.5	29.3	480	2 Q8GLN4	Q8gln4 pseudomonas
11	960	28.7	627	16 Q97PB8	Q97pb8 streptococc
12	957	28.6	617	16 Q8Y904	Q8y904 listeria mo
13	956.5	28.6	480	2 Q9F8X3	Q9f8x3 pseudomonas
14	956	28.6	627	16 Q8DMS8	Q8dms8 streptococc
15	951.5	28.5	627	16 Q8NZJ7	Q8nzj7 streptococc
16	949	28.4	655	16 Q97NW9	Q97nw9 streptococc

17	937.5	28.1	627	16 Q8K5Z4	Q8k5z4 streptococc
18	935.5	28.0	620	16 Q99Y91	Q99y91 streptococc
19	934	27.9	705	16 Q8DNI6	Q8dni6 streptococc
20	933.5	27.9	475	16 Q99WC9	Q99wc9 staphylococ
21	933	27.9	674	16 Q8K5K9	Q8k5k9 streptococc
22	931	27.9	676	16 Q8E208	Q8e208 streptococc
23	930.5	27.8	630	16 Q8G19	Q8g19 bacillus ha
24	929.5	27.8	475	16 Q8N12	Q8n12 staphylococ
25	929.5	27.8	636	16 Q9KF90	Q9kfi90 bacillus ha
26	927.5	27.8	655	16 Q8KST1	Q8kst1 streptococc
27	927	27.7	633	16 Q92FS7	Q92fs7 listeria in
28	924.5	27.7	634	16 Q8YAT6	Q8yat6 listeria mo
29	924	27.6	676	16 Q8E7G5	Q8e7g5 streptococc
30	921	27.6	674	16 Q8N240	Q8n240 streptococc
31	920	27.5	674	16 Q99XQ4	Q99xq4 streptococc
32	897.5	26.9	644	16 Q8DUF2	Q8duf2 streptococc
33	880.5	26.3	639	16 Q8DY04	Q8dy04 streptococc
34	876.5	26.2	639	16 Q8E3M3	Q8e3m3 streptococc
35	868.5	26.0	644	2 Q9KJ80	Q9kj80 streptococc
36	862	25.8	621	2 Q48408	Q48408 klebsiella
37	860.5	25.7	628	16 Q97J79	Q97j79 clostridium
38	859	25.7	617	16 Q8Y3R1	Q8y3r1 listeria mo
39	854.5	25.6	577	2 Q9L461	Q9l461 lactobacill
40	829	24.8	618	16 Q8Y882	Q8y882 listeria mo
41	828	24.8	465	16 Q8XQ82	Q8xq82 ralstonia s
42	824.5	24.7	649	16 Q8FB01	Q8f8b1 escherichia
43	822.5	24.6	636	16 Q9CFK9	Q9cfk9 lactococcus
44	813.5	24.3	616	2 Q46129	Q46129 clostridium
45	808	24.2	457	2 Q9F499	Q9f499 erwinia amy

## ALIGNMENTS

### RESULT 1

Q8NMD6 ID Q8NMD6 PRELIMINARY; PRT: 661 AA.  
AC Q8NMD6;  
DT 01-OCT-2002 (TREMREL. 22, Created)  
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)  
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)  
DE Phosphotransferase system IIC components,  
DE glucose/maltose/N-acetylglucosamine-specific (EC 2.7.1.69).  
GN CGL2642.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;  
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF005282; BAC00036.1; .  
DR InterPro: IPR001127; PTS\_EIIA.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR InterPro: IPR003358; PTS\_EIIA\_1; .  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD002243; PTS\_EIIA; 1.  
DR TIGRfams: TIGR00830; PTBA; 1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.  
KW Transferase, Complete proteome.  
SQ SEQUENCE 661 AA; 69148 MW; COE5E1C7833B9F74 CRC64;

Query Match 99.0%; Score 3310; DB 16; Length 661;  
Best Local Similarity 98.9%; Pred. No. 1.4e-188;  
Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MDHKLDAQRILRDIGEDNIYAAAHGATRLRLVLTQDKVDQSLDDDDPLKCTFTTGM 60  
|||||

Db 1 MDHKLQAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60  
QY 61 FOIIVPGDGDVHVFKELDGATSKDIIVSTEQKLVYVANNANWFSRAVKVLADIFVPLPI 120  
Db 61 FOIIVPGDGDVHVFKELDGATSKDIIVSTEQKLVYVANNANWFSRAVKVLADIFVPLPI 120  
QY 121 LVGGGLLMAINNVVAODLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFATK 180  
Db 121 LVGGGLLMAINNVVAODLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFATK 180  
QY 181 RFGGNEFLGAGIGMAVFPVTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTGTVPLV 240  
Db 181 RFGGNEFLGAGIGMAVFPVTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTGTVPLV 240  
QY 241 VSWILATIEKFLHKLRLMGADFLITPVLTLTLTGLTFTIAGPAMRWGDLIAHGLQGLY 300  
Db 241 VSWILATIEKFLHKLRLMGADFLITPVLTLTLTGLTFTIAGPAMRWGDLIAHGLQGLY 300  
QY 301 DFGGPGVGLLFGVLVSPVITGLHQSFPPIELFNFQGGSFIFATASMANIAQGAACLA 360  
Db 301 DFGGPGVGLLFGVLVSPVITGLHQSFPPIELFNFQGGSFIFATASMANIAQGAACLA 360  
QY 361 FFLAKSEKLGAGASGSAVIGITEPAIFGVNLRRLRWPFYIGITAGGALIALFDIK 420  
Db 361 FFLAKSEKLGAGASGSAVIGITEPAIFGVNLRRLRWPFYIGITAGGALIALFDIK 420  
QY 421 AVALGAAGFLGVVSDIDAPDMWFLVCVVTFVIAFGAAIAYGLXLYRRNGSIDPDATAP 480  
Db 421 AVALGAAGFLGVVSDIDAPDMWFLVCVVTFVIAFGAAIAYGLXLYRRNGSIDPDATAP 480  
QY 481 VPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSVSDAMFASKGLSGVAIVPTKGOL 540  
Db 481 VPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSVSDAMFASKGLSGVAIVPTKGOL 540  
QY 541 VSPVSKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLGTHFNPLKQGDVKA 600  
Db 541 VSPVSKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLGTHFNPLKQGDVKA 600  
QY 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPVNTYGLGEIAGANLLNVAKKEAVPAT 660  
Db 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPVNTYGLGEIAGANLLNVAKKEAVPAT 660  
QY 661 P 661  
Db 661 P 661

## RESULT 2

Q9S6S6 PRELIMINARY; PRT; 650 AA.  
ID Q9S6S6 AC Q9S6S6  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Enzyme II sucrose protein (EC 2.7.1.69).  
GN SACH.  
OS Lactococcus lactis.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1358;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=99173919; PubMed=10074089;  
Luesink E.J., Marugg J.D., Kuipers O.P., De Vos W.M.;  
"Characterization of the divergent sacB and sacA operons, involved  
in sucrose utilization by Lactococcus lactis.";  
J. Bacteriol. 181:1924-1926(1999).  
DR EMBL; Z97015; CAB09690.1; -  
DR HSSP; P45618; 2GPR.  
DR InterPro; IPR001127; PTS\_EIIB.  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF003358; PTS\_EIIB\_1; 1.  
DR Pfam; PF003367; PTS\_EIIB; 1.

DR Pfam; PF02378; PTS\_EIIC; 1.  
DR PRODOM; PD002243; PTS\_EIIB; 1.  
DR PRODOM; PD001476; PTS\_EIIB; 1.  
DR TIGRFAMS; TIGR00826; EIB\_glc; 1.  
DR TIGRFAMS; TIGR00830; PTBA; 1.  
DR PROSITE; PS00371; PTS\_EIIB\_1; 1.  
DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
KW Transferase.  
SQ SEQUENCE 650 AA; 69636 MW; 250E26F7664D4204 CRC64;  
Query Match 45.1%; Score 1508; DB 2; Length 650;  
Best Local Similarity 46.9%; Pred. No. 1.5e-81;  
Matches 313; Conservative 123; Mismatches 192; Indels 40; Gaps 10;  
QY 1 MDHKLQAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60  
Db 1 MNHQAVERILNAV-GRDNTQGAHCAATRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 59  
QY 61 FOIIVPGDGDVHVFKELDGATSKDIIVSTEQKLVYVANNANWFSRAVKVLADIFVPLI 118  
Db 60 YQIIVPGDVTYVEEFIKLTIGISEA-STADLKEIAGSKQKQNFVMAVLLSLDIFVPLI 118  
QY 119 PILVGGGLLMAINNVVAODLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFAT 178  
Db 119 PALVAGGLLMAINNVVAODLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFAT 178  
QY 179 TKRFGGNEFLGAGIGMAVFPVTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTGTV 238  
Db 179 TKRFGGNEFLGAGIGMAVFPVTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTGTV 238  
QY 239 LVSWILATIEKFLHKLRLMGADFLITPVLTLTLTGLTFTIAGPAMRWGDLIAHGLQGL 298  
Db 239 LGVAFILAKLRFHFKYLNDAIDFTPLSLVITGLTFTIAGPAMRWGDLIAHGLQGL 298  
QY 299 LYDFGGPGVGLLFGVLVSPVITGLHQSFPPIELFNFQ-----GGSFIFATASMANIA 352  
Db 299 LYNTLALGMLVFGSFYSAIVVTGLHQSFPPIELFNFQ-----GGSFIFATASMANIA 358  
QY 353 QGAACLAFFVFLAKSEKLGAGASGSAVIGITEPAIFGVNLRRLRWPFYIGITAGGALIA 412  
Db 359 QAGATFALLFTVTKIKTKALAAPAGVSAIGLITEPAIFGVNLRRLRWPFYIGITAGGALIA 418  
QY 413 LIALFDIKAVALGAAGFLGVVSDIDAPDMWFLVCVVTFVIAFGAAIAYGLXLYRRNGSI 472  
Db 419 FMGLFHLVAVLSGSAIGLIF-SIRAGYNLQFWSIFISFLIAFVVTSTYGRMEAK--SI 476  
QY 473 DPDATAAPVPAAGTTKAEAPAEFSNDSTIIQ-----APLTGEATLSVSDAMFAS 524  
Db 477 TKE-----KNQONATIQOPEKVIIDPVKSGELLAPINGFVPLSDVSDPVFSK 525  
QY 525 GKLGGSAIVPTKGOLVSPVSKIVVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLN 584  
Db 526 EIMGKGIKPKSGELFSPADGEIIIAVETGHAYCIKTK----NGGEVLLHIGIDTVSMN 581  
QY 585 GTHFNPLKQGDVKAAGLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPVNTYGLGEIAGAN 641  
Db 582 GNGFTQNVKVGOKVAGDLGSLGDFKEIKKSGLDVTIVITNSASYNEIPLSEN--VD 639  
QY 642 IEAGANLL 649  
Db 640 IKVGEKIL 647

## RESULT 3

Q93ML1 PRELIMINARY; PRT; 534 AA.  
ID Q93ML1 AC Q93ML1  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Sucrose-specific enzyme II of the PTS (Fragment).  
GN SCRA.  
OS Lactobacillus sakei.

OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;	
OC	Lactobacillus	
OX	NCBI_TaxID=1599;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Dudez A.-M., Chailiou S., Hissler L., Stentz R., Champomier-Verges M.,	
RA	Alpert C.-A., Zagorec M.,	
RT	"Physical and genetic map of the Lactobacillus sakei 23K chromosome.";	
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF010146; AAK92528.1; -	
DR	InterPro; IPR001127; PTS_EIIC.	
DR	InterPro; IPR003352; PTS_EIIC.	
DR	Pfam; PF00358; PTS_EIIC.1; 1.	
DR	Pfam; PF02378; PTS_EIIC.1; 1.	
DR	ProDom; PD002243; PTS_EIIC.1; 1.	
DR	TIGRFAMS; TIGR00830; PTBA; 1.	
DR	PROSITE; PS00371; PTS_EIIC.1; 1.	
FT	NON-TER	
SQ	SEQUENCE 534 AA; 56230 MW; 01804F9DE70C0089 CRC64;	
	Query Match 37.8%; Score 1264; DB 2; Length 534;	
	Best Local Similarity 47.6%; Pred. No. 3.6e-67;	
	Matches 265; Conservative 99; Mismatches 155; Indels 38; Gaps 10;	
QY	114 FVPLIPILVGGGLLMAINNVLAQDLFGPSLVEMFPQISGVAEMINLMASAPFAFLPVL 173	
DB	1 FVPLIPALTAGGLLMAINNVLTGGLEGASIVQMFPQWGFRAEIVNMWSAPFTLPIL 60	
QY	174 VGFATKRFNGNEELGAGIGMAMVFTLVNGDYVAATMTAGEMPMSLFLGLDVAQAGYQG 233	
DB	61 IAFSATKRFNGNPGVLAAGAMLVNPNLVNGYGVAESIATGHTMYHVFGLNTAQAGYQG 120	
QY	234 TVLPVLYSVTLATIEKFLKRLMGTADELITPVLTLGLGFTLFIAGPAMRWVGDLLA 293	
DB	121 QVPIPVGVAFTLANLEKFFHKLNDADVFTFPMLSIIITGLFTLVGPALRIVSNGVT 180	
QY	294 HGLQGLDFGPGVGLLGLVYSPITVITGLHQSPPIELELF----NOGGSFIFATASMA 349	
DB	181 DSLWVAYQTLGAVGMGIFGLYSAILVTLGLHQSPATETLLADIATKGSFIFPVAAMA 240	
QY	350 NIAQGAACLAFFFLAKSEKLGAGAGSVSAVLGITEPAFGVNLRLRWPFYIGIGTAI 409	
DB	241 NIAQGAATFAVFTTKNRQKSLTTSAGISAMLGITFPAUGVNLKLFPPFGLIASGI 300	
QY	410 GGATIALPDIKAVAGAGFLGVVSIDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRN 469	
DB	301 SSFIIGLHLVSVSNGPAGIIGFTAIAPKSPSEMGAIISFVAFVGTVLYGKKAMK-- 358	
QY	470 GSIDPDATAAPVPAGTAKAER--EAPA-----EFSNDSTIIQAPLTGEATLSVSDAMP 522	
DB	359 -----TTEEEIINEAPATPEYVERLQDEK--ISAPVTGRIVDLASVPDPVF 402	
QY	523 ASGKLGVAIVPTKGLQVSPVSGKIVVAFPSKAVFRTKAEKDSNVNIDILMHIGFDTVN 582	
DB	403 ASEANGKGIAMPTSDQVLAFTVGTITANTGHAYGI--KSDDGA--EVLHIGLDTVN 458	
QY	583 LNGTHFNPLKQGGDEVKAGELLCFDDAIIKAAGYEVVTPIVNSN---YKKTGPVNTYGL 639	
DB	459 LNCIGFEKIVQGGQVHSEGLLGHFDIDIKQAGLITPLMTIVINTAGYAGVDPLLTVDK 518	
QY	640 GEIEAGANLNV-AKKE 655	
DB	519 AAWQ-GEETIQLHAKKD 534	
RESULT 4		
Q8CN82	PRELIMINARY; PRT; 481 AA.	
ID	Q8CN82	
AC	Q8CN82	
DT	01-MAR-2003 (Tremblrel. 23, Created)	
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)	
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)	
DE	PTS system sucrose-specific IIBC component.	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_TaxID=1282;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Dudez A.-M., Chailiou S., Hissler L., Stentz R., Champomier-Verges M.,	
RA	Alpert C.-A., Zagorec M.,	
RT	"Physical and genetic map of the Lactobacillus sakei 23K chromosome.";	
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF010146; AAK92528.1; -	
DR	InterPro; IPR001127; PTS_EIIC.	
DR	InterPro; IPR003352; PTS_EIIC.	
DR	Pfam; PF00358; PTS_EIIC.1; 1.	
DR	Pfam; PF02378; PTS_EIIC.1; 1.	
DR	ProDom; PD002243; PTS_EIIC.1; 1.	
DR	TIGRFAMS; TIGR00830; PTBA; 1.	
DR	PROSITE; PS00371; PTS_EIIC.1; 1.	
FT	NON-TER	
SQ	SEQUENCE 534 AA; 56230 MW; 01804F9DE70C0089 CRC64;	
	Query Match 35.5%; Score 1186; DB 16; Length 481;	
	Best Local Similarity 49.9%; Pred. No. 1.4e-62;	
	Matches 235; Conservative 95; Mismatches 123; Indels 18; Gaps 6;	
QY	1 MDKDLAORILDRIGGEDNIVAAACATRLRLVLKDTKDVROSLDDDDPDLKGTETGGM 60	
DB	1 MSYKSAEELKAIGEENLDMACHATRLRLVLNDESKYDEDTLSNMDVVKGTETGGM 60	
QY	61 FOIIVGPGVDVHFVKELDDATSKDIADVSTEQKLVVANNANWFSRAVKVLADIFVPLPI 120	
DB	61 YQIIIGSGTVNKVFNELEKITGKE-ASTTSEVKDKSKHNPQKFKMLSDIFVPIIPA 119	
QY	121 LVGGGLLMAINNVLAQDLF-GPOSILVEMFPQISGVAEMINLMASAPFAFLPVLGFTAT 179	
DB	120 IVAGGLLMLNLIYTKDLFYDGKSLIDVHVSQFSLADNINIFANAPFTLLILIGFSA 179	
QY	180 KRFSGNEFLGAGIGMAMVFTLVNGDYVAATMTAGE-MPMWSLFLGLDVAQAGYQGVLPV 238	
DB	180 KRFSGNPLGALMLVHPGLMSADYDFPKALEEKGAIHPHWDVFGHLHINEVYQGVLEP 239	
QY	239 LVYSWTLATIEKFLKRLMGTADELITPVLTLGLGFTLFIAGPAMRWVGDLLAHGLOG 298	
DB	240 LVATYTLATIEKFLKRLMGTADELITPVLTLGLGFTLFIAGPAMRWVGDLLAHGLOG 299	
QY	299 LYDFGPGVGLLGLVYSPITVITGLHQSPPIELELF----NOGGSFIFATASMANIAG 354	
DB	300 LYDFGPGVGLLGLVYSPITVITGMHSTIATVETTLIADATKTGGSFIFPIATMSNIAOG 359	
QY	355 AACLAFFFLAK-SEKLGAGAGSVSAVLGITEPAFGVNLRLRWPFYIGIGTAIAGGAL 413	
DB	360 GAALAAFFIQLKQKLGAGSVSAVLGITEPAFGVNLRLRWPFYIGIGTAIAGGAL 413	
QY	414 IALFDIKAVAGAGFLGVVSIDAPD-----MMVFLVCAVTFVIA 454	
DB	420 ISFEKVAIATGATAGLPGFISINPTHAGWLHLYLGLMIAFVSVVTVLVLS 470	
RESULT 5		
Q8NV35	PRELIMINARY; PRT; 480 AA.	
ID	Q8NV35	
AC	Q8NV35	
DT	01-OCT-2002 (Tremblrel. 22, Created)	
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)	
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)	
DE	PTS system sucrose-specific IIBC component.	
GN	SCRA OR MW2299.	
OS	Staphylococcus aureus (strain MW2).	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_TaxID=196620;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,	
RA	Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,	
RA	Yamamoto K., Hiramatsu K.	
RT	"Genome and virulence determinants of high virulence community-	
RT	acquired MRSA.";	
RL	Lancet 359:1819-1827(2002).	
DR	EMBL; AP004830; BAB96164.1; -	

```
DR InterPro: IPR001996; PTS_EIIB.
DR EMBL: AP003365; BAB58539.1; -.
DR Pfam: PF00367; PTS_EIIB; 1.
DR InterPro: IPR001996; PTS_EIIB.
DR Pfam: PF00367; PTS_EIIB; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRFAMs: TIGR00826; EIB_glc; 1.
DR PROSITE: PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51218 MW; B3EBE81C08CODEDA CRC64;

Query Match 35.4%; Score 1183; DB 16; Length 480;
Best Local Similarity 50.4%; Pred. No. 2e-62;
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;

QY 1 MDHKLDAQRILRDIGEDNIVAAACATRLRLVLDKTDVDRQSLDDDDPLKGFETGGM 60
DB 1 MNYKQSAEILNAGSEENLDAMAHACATRLRLVLDNDESLVNEEALNNMNVVKGFTSGGQ 60
QY 61 FOIIVGPDVDHVFKELDATSKDIAVSTEQKLDVVANNANWFSRAVKVLADIFVPLPI 120
DB 61 YQIIIGSVNVKVFSELEKLTGKE-ASTTSEVKAQSAKNNPLQRFVKMLSDIFVPIIPA 119
QY 121 LVGGGLMAINNVVAODL-FGQSLVEMFPQISGVAEMINLMASAPFAFLPVLGFTAT 179
DB 120 IVAGGLMLGNLITAKDLFFSGKSLIDVYSQFAGLAEMINVFANAPFTLLPILIGFSAA 179
QY 180 KRFGGNEFTGAGIGMAMVPTLVNGYDVAATMTAGE-MPMWSLFGDLVAAQGYQGTVLVP 238
DB 180 KRFGGNFFLGAALGMILVHPSLMSAYDFPKAVEAGKAIPYWDVFGHLINQVGYQGVLP 239
QY 239 LVVSWILATEKFLHKLMTGADFLITPVLTLLTGLTFTAIQPMRWVGDLLAHGLQ 298
DB 240 LVAAVILASTEGLRKYPTVLDNLLTFLSIFITAFITFVFGVPIITRQLGYWLSGLTW 299
QY 299 LYDFGPGVGLLFGVYSPVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQ 354
DB 300 LYDFGGAIGLIFGLYAPVITVGHHSFIATVETLLIADATKIGSFIFPIATNSVAQ 359
QY 355 AACLAVFFLAK-SEKLGAGAGSVAVLIGTEPAIFGVNLRNRPFYIGTAAIGGAL 413
DB 360 GAAIAFAFFIKQNKKLGVASAAGISALLGITEPAMFGVNLKLRVPFIIAIVSGSIG 419
QY 414 IALFDIKAVAGLGAAGLVGVSDA--PDMVMFLCAVVTFTAFGAIAVGLYLVNRNGS 471
DB 420 IAFKVKAIATAGTAGLPGFISINPVHAGLHVFVGMVTISFII-----AIVTVLLSKRKN 475

RESULT 6
Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE PTS system, sucrose-specific IIBC component.
GN SCRA OR SAV2377 OR SA2167.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
SEQUENCE FROM N.A.
SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RA MEDLINE=21311952; PubMed=11418146;
EX Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshino A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
```

```
RL Lancet 357:1225-1240 (2001).
DR EMBL: AP003365; BAB58539.1; -.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRFAMs: TIGR00826; EIB_glc; 1.
DR PROSITE: PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;

Query Match 35.2%; Score 1178; DB 16; Length 480;
Best Local Similarity 50.4%; Pred. No. 4e-62;
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;

QY 1 MDHKLDAQRILRDIGEDNIVAAACATRLRLVLDKTDVDRQSLDDDDPLKGFETGGM 60
DB 1 MNYKQSAEILNAGSEENLDAMAHACATRLRLVLDNDESLVNEEALNNMNVVKGFTSGGQ 60
QY 61 FOIIVGPDVDHVFKELDATSKDIAVSTEQKLDVVANNANWFSRAVKVLADIFVPLPI 120
DB 61 YQIIIGSVNVKVFSELEKLTGKE-ASTTSEVKAQSAKNNPLQRFVKMLSDIFVPIIPA 119
QY 121 LVGGGLMAINNVVAODL-FGQSLVEMFPQISGVAEMINLMASAPFAFLPVLGFTAT 179
DB 120 IVAGGLMLGNLITAKDLFFSGKSLIDVYSQFAGLAEMINVFANAPFTLLPILIGFSAA 179
QY 180 KRFGGNEFTGAGIGMAMVPTLVNGYDVAATMTAGE-MPMWSLFGDLVAAQGYQGTVLVP 238
DB 180 KRFGGNFFLGAALGMILVHPSLMSAYDFPKAVEAGKAIPYWDVFGHLINQVGYQGVLP 239
QY 239 LVVSWILATEKFLHKLMTGADFLITPVLTLLTGLTFTAIQPMRWVGDLLAHGLQ 298
DB 240 LVAAVILASTEGLRKYPTVLDNLLTFLSIFITAFITFVFGVPIITRQLGYWLSGLTW 299
QY 299 LYDFGPGVGLLFGVYSPVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQ 354
DB 300 LYDFGGAIGLIFGLYAPVITVGHHSFIATVETLLIADATKIGSFIFPIATNSVAQ 359
QY 355 AACLAVFFLAK-SEKLGAGAGSVAVLIGTEPAIFGVNLRNRPFYIGTAAIGGAL 413
DB 360 GAAIAFAFFIKQNKKLGVASAAGISALLGITEPAMFGVNLKLRVPFIIAIVSGSIG 419
QY 414 IALFDIKAVAGLGAAGLVGVSDA--PDMVMFLCAVVTFTAFGAIAVGLYLVNRNGS 471
DB 420 IAFKVKAIATAGTAGLPGFISINPVHAGLHVFVGMVTISFII-----AIVTVLLSKRKN 475

RESULT 7
Q9L8G6 PRELIMINARY; PRT; 627 AA.
AC Q9L8G6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Scra (Fusion: PTS system, beta-glucosidase specific IIBAC
DE component).
GN SCRA OR CAC0423.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
SEQUENCE FROM N.A.
STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=20391269; PubMed=10937490;
RA Tangney M., Mitchell W.J.;
RT "Analysis of a catabolic operon for sucrose transport and metabolism
in Clostridium acetobutylicum ATCC 824.";
RL J. Mol. Microbiol. Biotechnol. 2:71-80 (2000).
RN [2]
```

SEQUENCE FROM N.A.  
RC STRAIN-ATCC 824 / DSM 792 / VRM B-1787;  
RX MEDLINE-21359325; PubMed-11466286;  
RA Noelling J., Breton G., Omalchenko M.V., Makatova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.D., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL: AF205034; AAF35839.1; -;  
DR EMBL: AE007557; AAK78403.1; -;  
DR HSSP: P20166; IGPR.  
DR InterPro: IPR001127; PTS\_EIIB.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR InterPro: IPR001254; Ser.protease\_Try.  
DR Pfam: PF00358; PTS\_EIIB.1; 1.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD002243; PTS\_EIIB; 1.  
DR ProDom: PD001476; PTS\_EIIB; 1.  
DR TIGRFAMS: TIGR00826; EIIB\_glc; 1.  
DR TIGRFAMS: TIGR00830; PTBA; 1.  
DR PROSITE: PS00371; PTS\_EIIB.1; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
DR Complete proteome.  
KW Complete proteome.  
SQ SEQUENCE 627 AA; 67596 MW; B835AB0238FCA436 CRC64;

Query Match 32.7%; Score 1098.5; DB 16; Length 627;  
Best Local Similarity 36.8%; Pred. No. 3e-57;  
Matches 236; Conservative 122; Mismatches 243; Indels 41; Gaps 8;

QY 1 MDHKDLAQRILRDIGEDNIVAAAHCASTRRLVLRKTDVDRQSLDDDDPDLKGTETGCM 60  
DB 1 MDYPTAKDLTKLKGKXNVAHCASTRRLVLRKTDVDRQSLDDDDPDLKGTETGCM 60  
QY 61 FQIIVPGDGVHVFKEKLDATSKDIIVSTEQKLD---VVANNANMFSAVKVLADIFVPL 117  
DB 61 YQIITIGGSVNVKYKAVFEGT-----GISESLSDTKKAAKNNLFFERFARMLSNIFVPI 116  
QY 118 IPIVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGT 177  
DB 118 IPIVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGT 177  
QY 117 IPIVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGT 167  
DB 117 IPIVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGT 167  
QY 178 ATKRFEGNEFLGAGIMAVFPTLVNGYDVAATMTAGEMPMSLFGLDVAQAGYQGTVL 237  
DB 178 ATKRFEGNEFLGAGIMAVFPTLVNGYDVAATMTAGEMPMSLFGLDVAQAGYQGTVL 237  
QY 168 AAREFTNPNYLAALGAIMHPDLQNAW---TLGEGIKHTINIFGLNIGVGYQGTVL 223  
DB 168 AAREFTNPNYLAALGAIMHPDLQNAW---TLGEGIKHTINIFGLNIGVGYQGTVL 223  
QY 238 VLVSWSILATIEKFLHRLMGTADFLITPVLTLLTGLFTFAIGPAMRWVGLLAHGLQ 297  
DB 238 VLVSWSILATIEKFLHRLMGTADFLITPVLTLLTGLFTFAIGPAMRWVGLLAHGLQ 297  
QY 224 ILISVWVSVYIEKRLKIVPEALDILTPFTLMTITGFFAMVYVGGGRFVGEISLGLQ 283  
DB 224 ILISVWVSVYIEKRLKIVPEALDILTPFTLMTITGFFAMVYVGGGRFVGEISLGLQ 283  
QY 298 GLXDFGPGVGLLGLVYSPVITGLHOSFPPIELELFNOGG---SFIPATASMANIAQ 354  
DB 298 GLXDFGPGVGLLGLVYSPVITGLHOSFPPIELELFNOGG---SFIPATASMANIAQ 354  
QY 284 TLYNTTGFSGLVFGGLYSLIVITGIHSHFATEAGLLANPAHKNFLLPINSANVAQ 343  
DB 284 TLYNTTGFSGLVFGGLYSLIVITGIHSHFATEAGLLANPAHKNFLLPINSANVAQ 343  
QY 355 AACLAVFFLAKSEKLGAGASVAVLGITEPAIFGVNRLRWPFYVIGTAAIGALI 414  
DB 355 AACLAVFFLAKSEKLGAGASVAVLGITEPAIFGVNRLRWPFYVIGTAAIGALI 414  
QY 344 GAALAVYFKTRDKKWKSIAPSPSCLLGITTEPAIFGVNRLRWPFYVIGTAAIGALI 403  
DB 344 GAALAVYFKTRDKKWKSIAPSPSCLLGITTEPAIFGVNRLRWPFYVIGTAAIGALI 403  
QY 415 ALFDIKVALGAAGFLGVSDADPMVNFVCAVVTPIVAFGAIAVGLVLRNRSIDP 474  
DB 415 ALFDIKVALGAAGFLGVSDADPMVNFVCAVVTPIVAFGAIAVGLVLRNRSIDP 474  
QY 404 VFTKAMTAVGTGIPGIAIVKQGSFLNY-----LIAMILAFGAFIATMVLGKEETEE 459  
DB 404 VFTKAMTAVGTGIPGIAIVKQGSFLNY-----LIAMILAFGAFIATMVLGKEETEE 459  
QY 475 DATAAPYAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSVSDAMFASGKLGSGVAV 534  
DB 475 DATAAPYAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSVSDAMFASGKLGSGVAV 534  
QY 460 DLNKETVNDKIKVEEVES-----VVSVPNGKVVLKKNVPDKTFAEGLIGDGIGVD 509  
DB 460 DLNKETVNDKIKVEEVES-----VVSVPNGKVVLKKNVPDKTFAEGLIGDGIGVD 509  
QY 535 PTKQLVSPVSGKIVVAFPPSHAFVTRKAEKDGNSVDILMHIGDVTYNLNGHTNPKKQ 594  
DB 535 PTKQLVSPVSGKIVVAFPPSHAFVTRKAEKDGNSVDILMHIGDVTYNLNGHTNPKKQ 594  
QY 510 PEDGEVSPIDGTVVHFETKHAAMSK-----NGVEMLIHIGDITVKMEGNGFKSPIND 565  
DB 510 PEDGEVSPIDGTVVHFETKHAAMSK-----NGVEMLIHIGDITVKMEGNGFKSPIND 565

595 GDEVKAGELLCEFDIDAIRKAGYEVVTPPIVWSNYKKTGPVNT 636  
DB 566 GEEVKAGDKLIQDLVREKAVSPVILVITVTHEDMGFVNS 607

RESULT 8  
Q9KLT8 PRELIMINARY; PRT; 479 AA.

ID Q9KLT8  
AC Q9KLT8  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE PTS system, sucrose-specific IIBC component.  
GN VCA0653  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE-20405833; PubMed-10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae";  
RT Nature 406:477-483(2000).  
RL EMBL: AE004395; AAF96554.1; -;  
DR TIGR: VCA0653; -;  
DR InterPro: IPR004719; PTS\_IIC\_glc.  
DR InterPro: IPR001996; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD001476; PTS\_EIIB; 1.  
DR TIGRFAMS: TIGR00826; EIIB\_glc; 1.  
DR TIGRFAMS: TIGR00852; PTS\_glc; 1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.  
DR Complete proteome.  
KW Complete proteome.  
SQ SEQUENCE 479 AA; 50292 MW; DC0E129229957E3C CRC64;

Query Match 32.7%; Score 1093; DB 16; Length 479;  
Best Local Similarity 47.8%; Pred. No. 4.5e-57;  
Matches 222; Conservative 85; Mismatches 151; Indels 6; Gaps 3;

QY 1 MDHKDLAQRILRDIGEDNIVAAAHCASTRRLVLRKTDVDRQSLDDDDPDLKGTETGCM 60  
DB 1 MDYPTAKDLTKLKGKXNVAHCASTRRLVLRKTDVDRQSLDDDDPDLKGTETGCM 60  
QY 61 FQIIVPGDGVHVFKEKLDATSKDIIVSTEQKLDVVANNANMFSAVKVLADIFVPLPI 120  
DB 61 YQIITIGGSVNVKYKAVFEGT-----VEMSTNDVASAGAEKQNAQVAGLSDFVRIIPA 119  
QY 121 LVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGTAT 179  
DB 121 LVGGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGTAT 179  
QY 120 IVAGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGTAT 179  
DB 120 IVAGGLLMAINVLRQDLGPGOSLVEMFPOISGVAEMINLMASAPFAELPVLVGTAT 179  
QY 180 KRFGGNEFLGAGIMAVFPTLVNGYDVAATMTAGEMPMSLFGLDVAQAGYQGTVL 239  
DB 180 KRFGGNEFLGAGIMAVFPTLVNGYDVAATMTAGEMPMSLFGLDVAQAGYQGTVL 239  
QY 240 VVSWSILATIEKFLHRLMGTADFLITPVLTLLTGLFTFAIGPAMRWVGLLAHGLQ 299  
DB 240 VVSWSILATIEKFLHRLMGTADFLITPVLTLLTGLFTFAIGPAMRWVGLLAHGLQ 299  
QY 240 VSAVILAKIENGRLKIVPEALDILTPFTLMTITGFFAMVYVGGGRFVGEISLGLQ 283  
DB 240 VSAVILAKIENGRLKIVPEALDILTPFTLMTITGFFAMVYVGGGRFVGEISLGLQ 283  
QY 300 YDFGGVPGVGLLGLVYSPVITGLHOSFPPIELELFNOGG---SFIPATASMANIAQ 355  
DB 300 YDFGGVPGVGLLGLVYSPVITGLHOSFPPIELELFNOGG---SFIPATASMANIAQ 355  
QY 300 YDSAGVFGALFGFIVAPFVITGMHSHFATETQLLADIVTTGGTFFIPFIAMSNIAQGA 359  
DB 300 YDSAGVFGALFGFIVAPFVITGMHSHFATETQLLADIVTTGGTFFIPFIAMSNIAQGA 359

QY 356 AACLAFFFLAKSEKLAGASGVSAYLITEPAIFGVNLRWPFYIGTAAIGGALIA 415  
DB 360 AALAVGVMTKRLKGVAPSGVTALLGITEPAMEGVNLRYPPIAICGAALASAFIT 419  
QY 416 LFDIKAVAGAGFLGVVSDAPDMVMFLVCAVTFVAFGMAI 459  
DB 420 LFNKRAQALGAAGLPGIITSINPQIGYIMGMAISFAAFALT 463

## RESULT 9

Q9KAS1 Q8GLN4 PRELIMINARY; PRT; 470 AA.  
AC Q9KAS1;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE PTS system, trehalose-specific enzyme II, BC component.  
GN BH2216.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL: AP001514; BAB05935.1; -;  
DR HSSP: P05053; 1IBA.  
DR InterPro: IPR004719; PTSIIC\_glc.  
DR InterPro: IPR001936; PTS\_EIIB.  
DR InterPro: IPR003352; PTS\_EIIC.  
DR Pfam: PF00367; PTS\_EIIB; 1.  
DR Pfam: PF02378; PTS\_EIIC; 1.  
DR ProDom: PD001476; PTS\_EIIB; 1.  
DR TIGRfams: TIGR00826; EIB\_glc; 1.  
DR TIGRfams: TIGR00852; EIB\_glc; 1.  
DR PROSITE: PS01035; PTS\_EIIB\_CYS; 1.  
KW Complete proteome.  
SQ SEQUENCE 470 AA; 50185 MW; 3CE67B1E9650F5B9 CRC64;

Query Match 30.3%; Score 1013.5; DB 16; Length 470;  
Best Local Similarity 44.4%; Pred. No. 2.3e-52;  
Matches 204; Conservative 90; Mismatches 150; Indels 15; Gaps 4;

QY 3 HKDLAQRILRDIGDENIVAAHCAATRLRLVLDKTDVDRQSLDDDDPDLLKGFETGGMFQ 62  
DB 4 YKKEVNAIVEAIGGADNIQTATHCVTRLRFLRVLRDEKGVDOEKLESLDIVKGSFSTNGQFQ 63  
QY 63 IIVGPGDVHVFREL-----DQATSKDIATVSTEQLKDVVANNANWFSRAVKVLADIFVP 116  
DB 64 VIIGQGVTVKVKELVAETGIGATKEDV-----KDAAKNTVYQRAVKTADIFIP 116  
QY 117 LPIVLVGGLLMANNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVNG 175  
DB 117 ILPAIVTAGLGMGINILTAEIGFYDGASVYIDPQWADFAGLIANTAFVLEFLG 176  
QY 176 FTATKRGGNEFLGAGIGMAWVPTLVNGDVAAATWATAGEMPWMSLFLGDLVAAQYQGV 235  
DB 177 WSAVKRFGGSELGIVGLMLVHPDLLNANWYCAQALEGEIPTWNLFGLTIEQVGQGV 236  
QY 236 LPVLVSWILATIEKFLHKLMTADFLTPVTLTLTGLTFFIAGPAMRWGDLALH 295  
DB 237 LPILFSAWILAKIEIFURKRVPSIQLLVAVPALLITGFIATAAIGCTFTTIGNTV 296  
QY 296 LOGLYDFGGPVGGLLGLVSPVITGLHQSFPPIELEFNQ-GGSFIFATASWANTAOG 354  
DB 297 FTSIFAAPVLVGGFLYLIYAPLVVTVGMHHTFLAVDLQLLIGTIGGTFPLPILVLSIAOG 356

QY 355 AACLAFFFLAKSEKLAGASGVSAYLITEPAIFGVNLRWPFYIGTAAIGGALIA 414  
DB 357 SAALAMFATREKRLKGLSLSSAVSAYLITEPAMFGVNIRYKFPFVCAIISAIGGAFI 416  
QY 415 ALFDIKAVAGAGFLGVVSDAPDMVMFLVCAVTFV1 453  
DB 417 TVNGVLANSIGVGLPGIITSIQAGFGVGVFFIGWVIAPIL 455

## RESULT 10

Q8GLN4 Q8GLN4 PRELIMINARY; PRT; 480 AA.  
AC Q8GLN4;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE TREP.  
GN TREP.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O8r1-96;  
RA Mavrodi D.V., de la Fuente L., Mavrodi O.V., Wellner D.M.,  
RA Thomasow L.S.;  
RT "Trehalose utilization operon of biocontrol strain Pseudomonas  
RT fluorescens O8r1-96";  
RL submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY131006; AAN06918.1; -;  
SQ SEQUENCE 480 AA; 51311 MW; B1F632B63042A10F CRC64;

Query Match 29.3%; Score 979.5; DB 2; Length 480;  
Best Local Similarity 40.8%; Pred. No. 2.5e-50;  
Matches 193; Conservative 110; Mismatches 161; Indels 9; Gaps 7;

QY 2 HKDLAQRILRDIGDENIVAAHCAATRLRLVLDKTDVDRQSLDDDDPDLLKGFETGGMF 61  
DB 4 DYSKIASLQSLGVDNIEQAHCYTRLRSLKDAARVDSATLNQIDLKVGSEFTGGLF 63  
QY 62 QIIVGPGDVHVFRELDDATSKDIATVSTEQLKDVVANNANWFSRAVKVLADIFVPLIPIL 121  
DB 64 QVIVGPGVEKVAALRELTLGL-AAATIDVKGQAKGNGQRLVRVLSDFVFPILPAL 122  
QY 122 VGGGLLMANNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVNGFTATK 180  
DB 123 VIAGLLMGVNNLLGAKGMFIAGQTLDDAYPHLDGWSNLINLMANTSFFVFLPALYGSAAK 182  
QY 181 RFGGNEFLGAGIGMAWVPTLVNGVD---VAATWATAGEMPWMSLFG-LDVAQYQGVTVL 236  
DB 183 RFGGSEILGIVGLMLVHPDLLNANWYCAVAGLEGGSLPYDFIEGWFRIEKVYQGOIL 242  
QY 237 PVLVSWILATIEKFLHKLMTADFLTPVTLTLTGLTFFIAGPAMRWGDLALH 296  
DB 243 PILLAAVYVMSVIERLRAVFNALQILLVITPITVITGVLAALAIIGPVTRHGLILTEGM 302  
QY 297 QGLYDFGGPVGGLLGLVSPVITGLHQSFPPIELEFNQ-GGSFIFATASMANIAQGA 355  
DB 303 VALFDLAPVIGGAIFGLLYAPLVVTVGMHMTFLAVDLQLISTQGGTFTFPMVMSNLAQGS 362  
QY 356 AACLAFFFLAKSEKLAGASGVSAYLITEPAIFGVNLRWPFYIGTAAIGGALIA 415  
DB 363 AALAVFYWTRNARDKSMASTSAISAYFGITEPAMFGVNIRYKFPFVCAIISALGCIPLS 422  
QY 416 LFDIKAVAGAGFLGVVSDAPDMV-MFLYCAVTVFVIAFGAATAYGLVLR 467  
DB 423 LNKIKASIAIGVGLPGFISI-VPDIYIPMFVGVWVIAVSPFALTCALSMKIVR 474

RESULT 11  
Q97PB8

ID	Q97PB8	PRELIMINARY;	PRT;	627 AA.
AC	Q97PB8;			
DT	01-OCT-2001 (TREMELrel. 18, Created)			
DT	01-OCT-2001 (TREMELrel. 18, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	PTS system IIABC components.			
GN	SP1722.			
OS	Streptococcus pneumoniae.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1313;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TIQR4;			
RC	MEDLINE=2137209; PubMed=11463916;			
RA	Tetelin S., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,			
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,			
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,			
RA	Mayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,			
RA	Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,			
RA	McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,			
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C., C.M.,			
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,			
RT	"Complete genome sequence of a virulent isolate of Streptococcus			
RT	pneumoniae."			
RL	Science 293:498-506(2001).			
RL	EMBL: AE007465; AAK75799.1; -			
DR	TIQR; SPI722;			
DR	InterPro; IPR001127; PTS_EIIA.			
DR	InterPro; IPR001996; PTS_EIIB.			
DR	InterPro; IPR003352; PTS_EIIC.			
DR	Pfam; PF00358; PTS_EIIA_1; 1.			
DR	Pfam; PF00367; PTS_EIIB; 1.			
DR	Pfam; PF02378; PTS_EIIC; 1.			
DR	ProDom; PD002243; PTS_EIIA; 1.			
DR	ProDom; PD001476; PTS_EIIB; 1.			
DR	TIGRFAMS; TIGR00826; EIIB-g1c; 1.			
DR	TIGRFAMS; TIGR00830; PTBA; 1.			
DR	PROSITE; PS00371; PTS_EIIA_1; 1.			
DR	PROSITE; PS01035; PTS_EIIB_CYS; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 627 AA; 65683 MW; 9298CDD7B2E6B85F CRC64;			
Query Match	28.7%; Score 960; DB 16; Length 627;			
Best Local Similarity	33.3%; Pred. No. 5e-49;			
Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;				
QY	1 MDKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			
DB	1 MNQEIARKVIDALGGRNENSVAHCAATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			
QY	61 FOITVPGDGVHVEK---LDDATSKDIATVSTEQKLDVYVANNKFSRAVKVLADIFVPL 117			
DB	61 VOIIFGTGVNMYDEVVVLGLPTS-----SKDDMKAEVAKOGNWFQRAITFGDVFPI 115			
QY	118 IPILVGGGLMANNVLVAQILFGPSILVEMFPQISGVAENILMASAPFAFLPLVVGFT 177			
DB	116 IPVITATGLFVGRGLFNALM-----PLPGDFATYITQILDTAFIILPGLVWVS 165			
QY	178 ATKRGGEFFLAGIGMAWPTTLVNGVDVAATMTAGEMPWLSFGLDVAAGQGVLP 237			
DB	166 TFRVFGNPAVGIVLGMVLVSGSLFNWAVA---OGGEVTAMNFFG-IPVVVGLQGVLP 221			
QY	238 VLVVSKILATIEKFLKRLMGADFLITPLVLTLLTGLTFTAIQPMRWVGDLLAHGLQ 297			
DB	222 AFIIGVGNKFKKRVKVPVDIVLLVFPVTLVNSILGLFVIGPVHVVNYILIAIK 281			
QY	298 GLYDFGGPVGGLFLGLVSPVITVTLGHOSFPPIELFNQGGSTF-ATASMANTAQAA 356			
DB	282 AILSNPFGGLGFLGVHQLIVSGVHHFNLVQLLAADHANFNIAITAAATAQAA 341			
QY	357 CLAVFELAKSEKLGLAGASVAVLGIPTAIFGVNLRRLRPPFVIGTCTAIGGALIAL 416			

DB	342 TVAVGVTKNPKLKTALFPAALSAFLGITPAIFGVNLRFRKPFLLSLIAGIAGGLASI 401			
QY	417 FDIKRAVALGAAGFLGVVSIDAPDMV-----MFLYCAVVTFVIAFGAALAYGLYLR 468			
DB	402 -----LGLAGTNGIITIIIGTMLYVNGOLPOYLLMVAVSFALGALTVMFCY----- 449			
QY	469 NGSIDPDATAAPVPAGTTKAAE-APAEFSDNSTIIQAPLTGEGAIALSSVSDAMFASGKL 527			
DB	450 --EDEVDATAAKRAEVAEKEEVAAPALQNETLV--TPIVGDVVALADVNDVPVSSGAM 505			
QY	528 GSGVAIVTKQLYSPVSGKIIVAPPSGHAFVARTKAEDGNSVDILKMHIGFDITVNLNGTH 587			
DB	506 GOGIVVXPSQGVVYAPADAESVIAPTGHAFTGLKTR-----NGAEVLHVGIDITVSMNGDG 561			
QY	588 FNPLRKQGDVKKAGELLCEFFDIDAIIKAAAGYEVTTPIVSVN---YKTCGPVNTYGLGEIEA 644			
DB	562 PETKVAQGNKVKAGDVLGTGFSNKIAAAGLDOTTMTIVTNRGDYASVAPVAT---GSAVK 618			
QY	645 GANLLNV 651			
DB	619 GDAVIEV 625			
RESULT 12				
Q8Y904	PRELIMINARY;	PRT;	617 AA.	
AC	Q8Y904;			
DT	01-MAR-2002 (TREMELrel. 20, Created)			
DT	01-MAR-2002 (TREMELrel. 20, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	Hypothetical protein lmo0738.			
GN	Lmo0738.			
OS	Listeria monocytogenes.			
OC	Bacteria; Firmicutes; Bacilliales; Listeriaceae; Listeria.			
OX	NCBI_TaxID=1639;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAN=EGD-e / Serovar 1/2a;			
RC	MEDLINE=21537279; PubMed=11679669;			
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,			
RA	Baquerio F., Berche P., Bloeker H., Brandt P., Chakraborty T.,			
RA	Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,			
RA	Dammann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,			
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,			
RA	Gautier L., Goebel N., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,			
RA	Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,			
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,			
RA	Nordstok G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,			
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tisseret A.,			
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,			
RT	"Comparative genomics of Listeria species."			
RL	Science 294:849-852(2001).			
RL	EMBL: AL591976; CAC98816.1; -			
DR	ListliList; LMO00738;			
DR	InterPro; IPR001127; PTS_EIIA.			
DR	InterPro; IPR001996; PTS_EIIB.			
DR	InterPro; IPR003352; PTS_EIIC.			
DR	Pfam; PF00358; PTS_EIIA_1; 1.			
DR	Pfam; PF00367; PTS_EIIB; 1.			
DR	Pfam; PF02378; PTS_EIIC; 1.			
DR	ProDom; PD002243; PTS_EIIA; 1.			
DR	ProDom; PD001476; PTS_EIIB; 1.			
DR	TIGRFAMS; TIGR00826; EIIB-g1c; 1.			
DR	TIGRFAMS; TIGR00830; PTBA; 1.			
DR	PROSITE; PS00371; PTS_EIIA_1; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 617 AA; 65447 MW; E7144ACDA7846A03 CRC64;			
Query Match	28.6%; Score 957; DB 16; Length 617;			
Best Local Similarity	35.1%; Pred. No. 7.4e-49;			
Matches 225; Conservative 124; Mismatches 228; Indels 64; Gaps 17;				
QY	1 MDKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			

DB 1 MDYQKLAKELIANWGEENRVSWECAFLRFLKYNKEADKKQIESISGVISVVENAQQ 60  
QY 61 FOIIIVG--PGVDVHV---FKLEDDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIFV 115  
DB 61 LOVIIGNTVGDYVYKALGSFTKLTDDGSEIAGT--KD---SDGNFSLKALDVISGIFT 114  
QY 116 PLIPILVGGGLMAINNVVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLV 175  
DB 115 PILGALAGGGLKGLIMLT---FG-----WLTSSGTQOILVAAADSVFYFLPLILIA 165  
QY 176 FTAIRKRFEGNEFLGAGIGAMVFPITLVNGYDVAAMTAMTAGEMPMWSLFGLDVAQAQYQGTV 235  
DB 166 YTAARKEGANPPVAITAAGALVYPMINLFNSGAHITFLQIP-----VVLMSYSFSV 217  
QY 236 LPVLVSVTLATIEKFLHKLRLMGADFLITPVLTLTLGLFLTFIAIGPAMRWGDLIAHG 295  
DB 218 IPILAVWFLSIDERFLNSKIHEAAKFTLTPMCLMLIYPLFLAFGLPGLTFISOGLASG 277  
QY 296 LQGLYDFGPGVGLLFLGLVYSPITVITGLHQSPPPIEL-ELFNQGGSGFIFATASMANIAQG 354  
DB 278 YTFIYNLSPIVAGAFNGAQQWLVIFGIHWGFVPMINNLNLSRYGRDTMIAMVGPSFAQA 337  
QY 355 AACLAVFTFLAKSKLKLGLAGASVSAVLGITEPAIFGNLRLRWPFYIGIGTAAIGGALI 414  
DB 338 GASLGVFLTKKPEVKAIAGSAALTOFFGITEPSIYGTLYKKPFVVIASIAAGIAGAI 397  
QY 415 ALPDIKAVAGAGAGFLGVVSDAPDW---MFLVCAVTFVIAFAAATAY-----GLYLV 466  
DB 398 -----GAAGSSGAANA-IPGILTLPIFGKGFVGFIL--GLAVAYILSAIGTYFF 444  
QY 467 RRNGSIDPDAAPVPAAGTTKAEAPAFESNDSTIIQAPLGEATALSVDAMFASOK 526  
DB 445 --GYKDEKADGT--APTKEAKETGVE-----AEVIVSPIRGNIVPLNEVKDEAFSAGL 494  
QY 527 LGSVAIVPTKGLVSPVSKIVVAPPSGHAPAVRTKAEKDSNVDILMHIGDTVNLNGT 586  
DB 495 LKGVALVPOEGKLIISPVNCTIETAPTGHAGIIRS-----DKGVEILLHVHGEDTVLNGK 550  
QY 587 HFNPLKKQSGDEVKAGELLCFDDIDAKAGYEVTTPIVSN 627  
DB 551 YFKLLVAGQDRVLVGQALLFDELEAKADGYDITPIVVTN 591

## RESULT 13

ID Q9F8X3 PRELIMINARY; PRT; 480 AA.  
AC Q9F8X3;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE TREP.  
GN TREP.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17400;  
RX MEDLINE=98027372; PubMed=9361421;  
RA Gaballa A., Abeysinghe P.D., Ulrich G., Matthijs S., De Greve H.,  
RA Cornelis P., Koedam N.;  
RT "Trehalose induces antagonism towards Pythium debaryanum in  
RT pseudomonas fluorescens ATCC 17400.";  
RL Appl. Environ. Microbiol. 63:4340-4345(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17400;  
RX MEDLINE=21030915; PubMed=11191810;  
RA Matthijs S.L.C., Koedam N., Cornelis P., De Greve H.;  
RT "The trehalose operon of Pseudomonas fluorescens ATCC 17400.";  
RL Res. Microbiol. 151:845-851(2000).

DR EMEL; AF229829; AAC31031.1; -;  
DR InterPro; IPR001996; PTS\_EIIB.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR Pfam; PF00367; PTS\_EIIB; 1.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR ProDom; PD001476; PTS\_EIIB; 1.  
DR TIGRFAMS; TIGR00826; EIB\_glc; 1.  
DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
SQ SEQUENCE 480 AA; 51391 MW; 1FC73A88E302C013 CRC64;  
  
Query Match 28.6%; Score 956.5; DB 2; Length 480;  
Best Local Similarity 39.4%; Pred. No. 5.8e-49;  
Matches 186; Conservative 115; Mismatches 164; Indels 7; Gaps 5;  
  
QY 2 DKDLAQRIILDIGEDNIVAAAHCAATRLRLVLRKDVDRQSLDDDDPDLKGFETFGGMF 61  
DB 4 DYSNIAREILLEHSGSDNLEQAHCVTRLRLALKDPSLVNASALNQVDLVKGSFFETGLF 63  
QY 62 QIIVGPGDVHVFKELDSDATSKDIATVTEQLKDVVANNANWFSRAVKVLADIFVLPIL 121  
DB 64 QWVIGPGEVEKYYAALREQTGL-AAATIAVKKKGADKTNALQRLVRFSDVFMPIPAL 122  
QY 122 VGGGLMAINNVVAQDLF-CPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFATK 180  
DB 123 IIAAGLLGINNLGAKGMFIEGQILLLEAYPNLDGLWSLINLMANTSFFVLPAVGSAAK 182  
QY 181 RFGNEFIAGIGAMVFPITLVNGYD---VAATMTAGEMPWWSLFG-LDVAQAQYQGTVL 236  
DB 183 RFGSEILGIVGLMLVHPDNLNANWYKAVAGLDGOSLPYDFIFGWFQIEKVYQGOIL 242  
QY 237 PVLVSVTLATIEKFLHKLRLMGADFLITPVLTLTLGLTFIAGPAMRWGDLIAHGL 296  
DB 243 PILMAAYVMSVTEKWLRAVPNAQLLVPTTIVTIVGLALAIIGPVTRHLGILITEGV 302  
QY 297 QGLYDFGPGVGLLFLGLVYSPITVITGLHQSPPPIELFN-QGGSFFATASMANIAQGA 355  
DB 303 VTLFLAPWGAIFGLLYAPLITGMHMFAYLDQLISTOGGTFIPMTVMNLAQGS 362  
QY 356 ACLAVFEFLAKSEKLKLGLAGASVSAVLGITEPAIFGNLRLRWPFYIGIGTAAIGGALIA 415  
DB 363 AALGVYMSRNARDKSMASSTSAISAYFGITEPAMEGVNLRKFPFYAALIGLSGIFLS 422  
QY 416 LFDIKAVAGAGFLGVVSDAPDWMMFLVCAVTFVIAFAAATAYGLYLV 467  
DB 423 LNKVQASAIQVGGPLPGFISIIPIQYIPSPFVIGVIAIVPVFLTCGLSKMKIVR 474  
  
RESULT 14  
Q8DNS8  
ID Q8DNS8 PRELIMINARY; PRT; 627 AA.  
AC Q8DNS8;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Phosphotransferase system enzyme II (EC 2.7.1.69).  
GN SCRA OR SPRI566.  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21429245; PubMed=11544234;  
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burett S.,  
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicas T.I.,  
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,  
RA Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";



Db 567 VAVGOVRKGDLLGHFDPSPKIAEAGLDDTTMMIVTNSADYOSVDILAQGHVLIG 620

Search completed: October 1, 2003, 05:03:58  
Job time : 72 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 05:00:51 ; Search time 27 Seconds  
(without alignments)  
1035.932 Million cell updates/sec

Title: US-10-019-284A-2  
Perfect score: 3342  
Sequence: 1 MDHKDLAQRILRDIGEDNI.....TEAGANLLENVAKKEAVPATP 661

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents,AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	923	27.6	628	4	US-09-107-532A-5288
2	826	24.7	631	4	US-09-107-532A-3902
3	737	22.1	583	4	US-09-107-532A-6811
4	644	19.3	590	4	US-09-107-532A-6004
5	528.5	15.8	243	4	US-09-134-001C-4114
6	511	15.3	478	4	US-09-134-001C-5065
7	495.5	14.8	427	4	US-09-107-532A-5530
8	437.5	13.1	482	4	US-09-107-532A-6644
9	401	12.0	527	4	US-09-634-238-368
10	329	9.8	696	4	US-09-134-001C-4296
11	315	9.4	656	4	US-09-107-532A-5083
12	296.5	8.9	585	4	US-09-107-532A-4366
13	271.5	8.1	470	4	US-09-107-532A-5798
14	267	8.0	170	4	US-09-134-001C-3134
15	230	6.9	686	4	US-09-252-991A-18115
16	183.5	5.5	98	4	US-08-936-165A-419
17	181.5	5.4	153	4	US-09-107-532A-4365
18	172.5	5.2	355	4	US-09-107-532A-5592
19	170.5	5.1	655	4	US-09-134-001C-5531
20	168.5	5.0	661	4	US-09-134-001C-4303
21	161.5	4.8	479	4	US-09-107-532A-6243
22	151.5	4.5	490	4	US-09-134-001C-5116
23	147	4.4	470	4	US-09-328-352-5397
24	146.5	4.4	387	4	US-09-107-532A-6002
25	144.5	4.3	498	4	US-09-107-532A-5037
26	142.5	4.3	451	4	US-09-107-532A-5352
27	142.5	4.3	487	4	US-09-107-532A-6319

28	142	4.2	495	4	US-09-252-991A-19278	Sequence 19278, A
29	142	4.2	575	4	US-09-328-352-7529	Sequence 7529, Ap
30	140	4.2	481	4	US-09-107-532A-6912	Sequence 6912, Ap
31	137.5	4.1	505	4	US-09-328-352-7470	Sequence 7470, Ap
32	135.5	4.1	359	4	US-09-134-001C-3728	Sequence 3728, Ap
33	134	4.0	353	4	US-09-134-001C-4325	Sequence 4325, Ap
34	134	4.0	651	4	US-09-252-991A-18065	Sequence 18065, A
35	133	4.0	403	4	US-09-328-352-7791	Sequence 7791, Ap
36	128	3.8	731	4	US-09-340-736E-1	Sequence 1, Appl1
37	126.5	3.8	417	4	US-09-107-532A-5437	Sequence 5437, Ap
38	126.5	3.8	480	4	US-09-252-991A-17687	Sequence 17687, A
39	125.5	3.8	512	4	US-09-107-532A-32995	Sequence 32995, A
40	125.5	3.8	528	4	US-09-252-991A-6836	Sequence 6836, Ap
41	125	3.7	432	4	US-09-107-532A-6836	Sequence 1, Appl1
42	125	3.7	731	2	US-08-911-364-1	Sequence 2, Appl1
43	125	3.7	733	3	US-08-464-700-2	Sequence 4, Appl1
44	125	3.7	3519	3	US-09-428-517-4	Sequence 5006, Ap
45	122.5	3.7	533	4	US-09-107-532A-6006	

ALIGNMENTS

RESULT 1  
US-09-107-532A-5288  
; Sequence 5288, Application US/09107532A  
; Patent No. 6383275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESS: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 628 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...628  
SEQUENCE DESCRIPTION: SEQ ID NO: 5288:

US-09-107-532A-5288

```

: OPERATING SYSTEM: <Unknown>
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/107,532A
: FILING DATE: 30-Jun-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/085,598
: FILING DATE: 14 May 1998
: APPLICATION NUMBER: 60/051571
: FILING DATE: July 2, 1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Ariniello, Pamela Deneke
: REGISTRATION NUMBER: 40,489
: REFERENCE/DOCKET NUMBER: GTC-012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (781)893-5007
: TELEFAX: (781)893-8277
: INFORMATION FOR SEQ ID NO: 3902:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 631 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Enterococcus faecium
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (B) LOCATION 1...631
: SEQUENCE DESCRIPTION: SEQ ID NO: 3902:
US-09-107-532A-3902

Query Match          24.7%   Score 826;   DB 4;   Length 631;
Best Local Similarity 30.5%   Pred. No. 4e-76;
Matches 196;   Conservative 123;   Mismatches 243;   Indels 80;   Gaps 11;

QY      1 MDHKDLAQRILRIDIGEDNIVAAAHCAHTRLRVLVKOTKVDVDRQSLDDDDPLKGTFFETGGM 60
DB      28 MDYSQLAKDIVRFYGGSENSYNYHCAHTRLRFTLTKDNKKADKKEKVEQLEGVITVVEAGM 87

QY      61 FOITVPGDVHDVFKELDDATSKDIAVSTEQKQDVANNANWFSRA-----VKVLADIF 114
DB      88 FQVVVG-----NAVNEYVDVLSKQM-----KLEDDASSGKRGTEKKGLNLSFIDMIAAVF 137

QY      115 VPLIPIVLGGLLMAINNLVVAQDLFGPQSLVEMFPQISGVAEMINLMASPAFLPLV 174
DB      138 APTLGVLAGSLIKGVLAQCTSLNLTTE-----SGFYIILNAAADAFFYFLPLFL 188

QY      175 GFTATKRFEGNEFLGAGIGMAVPEPTLVNGYDVAAWTAGEMPWMSLFGLDVQAQGYGT 234
DB      189 AYTAARKENTDRFIAMVIAAALVPTIVSAYSDSITL-----RFLGMPVILARYTST 240

QY      235 VLPVLVSVWILATTEKELHKLMGTAQDLITPVLTLTLTGFLTFIATGPMRWGDLIAH 294
DB      241 VIPAILAVWLVSYTEPKIRKSLHESINRLTFFCIIVMVFPIILLVGPPIADYASQLIAA 300

QY      295 GLOGLYDFGPGVGGLLFGLVYSPVITIGLHQSFPPIELELFNQGSGSFIATASMANI-AQ 353
DB      301 GYLAVYNFSPVLSGAVIGGFQWVLFTGLHWGLVPVMTNLSFYGRDPLGPACMTAVAAQ 360

QY      354 GAACLAVFVFLAKSKELKLAGASGVSAVLGTEPAIFGNVLRLRWPFFYIGIGTAIGGAL 413
DB      361 AGAVLGVFLTKNKKVKSLSUSAFITALFGTEPAVYGVTLKYRPFYIACICGAIFGV 420

QY      414 IALFDIKAVALLGAAGFL-----GVVSDIDAPQVMVFLVCVVYTFVIAFGAAIYGLYL 465
DB      421 AGAAGAGALAVATRSILSFPIYIGEGFWLVASYFLAMISSCMILTFLGY----- 470

QY      466 VRRNGSIDPDATAAPVAGTTKAEAREAPESNDSTIIQAPLTGEATALLSSVSDAMPASG 525
DB      471 -----KDETE-----ESSKDIVLSTPAAGEIIDLSEVNDPTFASG 506

QY      526 KLGSGVAIVPTKGLVSPVSGKIVVAFPSGHAFVTRTKAEDGNSVDILMHIGFTVNLNG 581

```

```

Db      507 SLGEGFAIPTDGIYSSVNGEYSTVEPTKHAIGV--VSEGA--BILIHIGDITVNLNG 562
QY      586 THENPLKQGDVEYKAGELCEPDIDAIAKAAAGYEVVTPPIVYSN 627
Db      563 KYQSAVSDGKVKRGDOLLMEVDLQELIKEGYDPTTWIVTN 604

RESULT 3
US-09-107-532A-6811
; Sequence 6811, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6811:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...583
; SEQUENCE DESCRIPTION: SEQ ID NO: 6811:

US-09-107-532A-6811
Query Match      22.1%; Score 737; DB 4; Length 583;
Best Local Similarity 30.6%; Pred. No. 5.4e-67;
Matches 180; Conservative 119; Mismatches 239; Indels 50; Gaps 13;

QY      87 VSTOLKDVANNANWTSRAVKVLADIFVPLIPLVGGGLMAINNVLVA---QDLRGPQ 143
Db      7 VSTEGGAAAKQNLHPQVRAIVLAEITPLIPAIIVGGILGPRNVLEGIQFESLGG-- 64
QY      144 SLVMEFPOISGVAEMINLMASAPFAFLPVLVGFTATKRFSGNEFLGAGIGMAMVFEPLVN 203
Db      65 TIVSHKFWNGVNAFLWLPGBAIFHFLPVGVITWSIAKMGTTQILGVLGTLIVSPOLLN 124
QY      204 GYDVAATMTAGEMPWMSLGLDVAQAGYQGVTLVPLVYSWLATIERFLHKLRLMGTAFFL 263

```

```

Db      125 AYSVAST-AAADIPFWDGFAQVQMGVQAOVIPAMLAGFMLAYLEIFFRKIYIPOSISMI 183
QY      264 ITPVLTLLTGLTETIAIGP----AMRWVGDLALHAGLQGLYDFGPGVGLFLGLVYSPIV 319
Db      184 FVPLFSLPTVLAHVILSPIGTWGWSISNIVNTGLTSSIXW---LFSAVFEGELYAPLV 240
QY      320 ITGLHQSFPPIELELFNOGGSF-IFATASMANIAGAACIAYFFLAK-SEKLKGLAGASG 377
Db      241 ITGLHMMNAIDMQLIADFGSTNLMPIALNSIAQGSASVLAIIIFLHRNKKKEEOLISIPAM 300
QY      378 VSAVLGITEPAIFGVNLRWPFYIGIGTAAIGGALIALFDIKAVAGAACGFLGVVVSIDA 437
Db      301 ISCYLGVTPEAMFGLNXYVFPVAAVGVSGLAGMFANLMGVRAANICVGGGLPGLALQOA 360
QY      438 PMVMFLCAVVTFFVIAAGAAIAYGLYLVRNG---SIDP--DATAAPVPAGTTKAAEAA 492
Db      361 ETWVFFIIAMIIAIVPIEGLTIIF-----RRQGLNKIDPAVPPENAAADVQLQTANGATAT 415
QY      493 PAEFSNDSTIIQAPLTGEAIA-----LSSVSDAMFASGLGSGVAIVPTK 537
Db      416 PQSPEPVS-----TGTAVATKETLFAVAAGTKEITEVNDPVFSQKMGDGAIVEFSN 469
QY      538 GOLVSPVSGKIVVAPPSGHAFVARTKAEKGSNVDMHIGFTVNLNTHENPLKKQODE 597
Db      470 GKVPYVNGKVTSPETKHAIGILS-----NEGLEVLVHMGLDTVELKGVPNFVFKEGYL 525
QY      598 VKAGELLCEFDIDAIAKAAGYEVVTPPIVYSNKKTKGTVNTYGLGTEAG 645
Db      526 VTPETLIAEMDLPEIEQAGKKTDIIVALTNNKXVAGLSLDQSGLVRPQ 573

RESULT 4
US-09-107-532A-6004
; Sequence 6004 Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6004:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```
;
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...590
; SEQUENCE DESCRIPTION: SEQ ID NO: 6004:
US-09-107-532A-6004

Query Match          19.3%; Score 644; DB 4; Length 590;
Best Local Similarity 26.6%; Pred. No. 2.2e-57;
Matches 164; Conservative 130; Mismatches 249; Indels 74; Gaps 11;

QY 3 HKDLAQRILRDGGEDNIVAAACATRLRLVLDKDVDRSLDDPDLKGTFFGGMFQ 62
Db 24 YQELARKIVENYGGKENGINGLTHC:TRLRFKLKNEEKANDEILKMGDIVVMRAGGQY 83
QY 63 IIVGPDVDHVKELDDA--TSKDJAVSTEOLKDVVANNANWESRAVVLADIFVPLIPI 120
Db 84 VVIG-NHVPVVEEVIKAGNLTFDEAVTEKRP-----FMDLIDIGSCQFPFLAI 134
QY 121 LVGGGLMAINVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPPAFLPVLVGTATK 180
Db 135 LAAGMIRGLTAFLV-----ELGAFDRGSGTFVMDNIGDSVFQFMPVILGLTAAR 185
QY 181 RFGGNEFLGAGIGMAWVPTLVNGYDVAATMTAGEMPWMSLFGDLVAAQ----- 229
Db 186 KKYNEFVGMILGAALMNPSL----SLRLSAAEAPLTIIFSCTIFEAPIYQIVFGIPW 241
QY 230 ---GYOGTVLPVLVWSWILATIEKFLHKLMTGADFLITPVLTLTLTGLTFLTAIGPAMR 286
Db 242 IARNYASSVPIIFIVLLASQIQPKIKLVPEMIANFVFPTVLITMPLGFLLVGPVET 301
QY 287 WYGDLLAHGLOGLYDFGGVPGGLGLVLYSPVITVGLHOSFPPIELELFN-QGGSFIPAT 345
Db 302 FATDILMAGFELLALSPIYGAIVGFTWQIIVNFGHLHWAIVPMGLMDFSVNGWQNIWTP 361
QY 346 ASMANIAQGAACLAFFFLAKSKLKLGLAGAGSVAVLGTETPAIFGVNLRWMPYVIGIG 405
Db 362 VAVVSGTAAALTYALFKLRNPKRAIPAIVSGVIGITEPAIYGFTEPRKFIETCY 421
QY 406 TAAIGGALLATPDIKAVAGAGELGVVSI-----DAPDMVFLCAVAVTFVIAFGAIA 460
Db 422 GGAIGGAYGLMNLTSWNGGGIGFTIPNYIRPDGLDIDVINVLIGIAIAMVSTLTF 481
QY 461 YGLYLVRNGSIDPDATAAPVPAAGTTKABEAAPAEFSNDS--TIICAPLTGSAIALSSVS 518
Db 482 F-----WKDEAGETDDIQKSGKEIVKTPIOGQIAPLNAAK 517
QY 519 DAMFASGLGSGVAIVPTKGQVSPVSGKIVVAFPSGHAFAYVTRKAEDGNSVDILMHICF 578
Db 518 DAAFAOGTILGRGLIYPEKGEVRAPDGTIVMTLFTPKQAIMVSE----TGLELLIHVGL 573
QY 579 DTVNLNGTHFNPLKKOG 595
Db 574 DTVLECKYFESLVQOG 590

RESULT 5
US-09-134-001C-4114
; Sequence 4114, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```

```
;
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4114
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4114

Query Match          15.8%; Score 528.5; DB 4; Length 243;
Best Local Similarity 48.6%; Pred. No. 4.1e-46;
Matches 107; Conservative 43; Mismatches 55; Indels 15; Gaps 3;

QY 250 KFLHKLMTGADFLITPVLTLTLTGLTFTTAIGPAMRWGDLAHLAHLQGLYDFGGPVGGL 309
Db 13 KGLKFKVILTVLDNLTPLLSIFATIFLFGVPTVRLQGLWLSGLTWLYEFGGAIGGL 72
QY 310 LFLGVYSPVITVGLHOSFPPIELELF-----NQGSFIFATASMANIAQGAACLAFFFLAK 355
Db 73 IFGLLYAPIVITGMHHIFIAVETTLIADATKTGGSFTPIATMSNIAQGAALAAFFIHK 132
QY 366 -SEKLKLAGASGVSAVLGITETPAIFGVNLRWMPYVIGITAAIGGALLATFDIKAVAL 424
Db 133 QNKKLKGVAAGSALSILGITEPAMFGVGNLKLRYPPFICAVAGSGIGNAYISFFKVAIAL 192
QY 425 GAAGFLGVSIDAPD-----MYMFLVCAVAVTFVIA 454
Db 193 GTAGLPGFISINPTHAGLWHLHYLIGMLIAFVSVVVVLVLS 232

RESULT 6
US-09-134-001C-5065
; Sequence 5065, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5065
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5065

Query Match          15.3%; Score 511; DB 4; Length 478;
Best Local Similarity 29.8%; Pred. No. 8.3e-44;
Matches 147; Conservative 90; Mismatches 199; Indels 58; Gaps 16;

QY 6 LAQRILRDIGGEDNIVAAAHCAHCAHCAHCAHCAHCAHCAHCAHCAHCAHCAHCAHCAH 65
Db 11 LAKDITHALGGSQNSIINIIHCHMTRVRIKVNDAKVNVDKLSINGVLGVVE-DERIQVW 69
QY 66 GPGVDVHFKELDDATSKDIAVSTEQLKDVVA-----NNANWESRAVK 108
Db 70 GPGIVNKVAKLMADQSGATLAETTESQSYKQAEKRAYEHKKFOFQSRQSKM-NKVLK 128
QY 109 VLADIFVPLIPLVGV-----GGLLMAINVLVAQDLFGPQSLVEMFPQISGVAEMINLMAS 164
Db 129 STANIFILIPAFIAGLUGIAGIAILSNLLTAGSISG-----QWITQIVTVLVNLIK 180
QY 165 APFAFLPVLVGTATKTRPGGNEFELGAGIGMAWVPTLVNGYDVAATMTAGEMPWMSLFL 224
Db 181 GMLFYLAIFTGINSKAVFEGATPGLGVIGGT-----TLTG-----ITDENPKNIFTG 229
QY 225 DVAQAGYOGTVLPVLVWSWILATIEKFLHKLMTGADFLITPVLTLTLTGLTFLTAIGPA 284
Db 230 EHLAG-OGGIIGVIFAFWLLSNVSKRLHKLIPNSIDIIVTPTITLLIGLITFIIMPL 288
```

```

QY 285 MRWGDLLAHGLOGVDFGPGVGLLFLGLVSPVITGLHQSPFPPELEFNCGGS-FIF 343
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 AGFVSGLYVINWIGVGIFSGFPIGAFPLVLMGLHHFTPHIELINQSTYLL 348
QY 344 ATASMANIQAAGACLAVFF-LAKSEKLG-LAGASCVSAVLGTEPAIFGVNLRWPPY 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
349 PIAAMAGAGQVGAIALWRCRCKNELNTLKGALPV-GFLGTEPLYGVTLPGRPF 407
QY 402 I-----GICFAATGGALIALFDIKAVAGAGFLGVVSDADPMVM-FLVCVVVTFVIAF 455
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 TACIGGGVGAIVGG-----IGHGATAVGPGS-ISLLPLIANNMGLYIVGLLAAYTGGF 462
QY 456 GAAIAYGLYLVRN 469
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
463 IFYFEGTTKEMN 476

RESULT 7
US-09-107-532A-5530
; Sequence 5530, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5530:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 427 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...427
; SEQUENCE DESCRIPTION: SEQ ID NO: 5530:
US-09-107-532A-5530

```

```

Query Match 14.8%; Score 495.5; DB 4; Length 427;
Best Local Similarity 29.3%; Pred. No. 2.7e-42;
Matches 123; Conservative 88; Mismatches 192; Indels 17; Gaps 7;

```

```

QY 238 VLVSWILATIEKFLHKLKMGTAFLITPVLTLLTGLTFEIAIGPAMRWYDGLLAHGLQ 297
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
17 VICAWLLSIVEKKLHKVVPDAVDIIVTPTLSLIAIGLIEIFFIMPLAGFISDGMVGGIN 76
QY 298 GLYDFGGVGGLLGLGLVSPVITGLHQSPFPPELEFNQ-GGSFIFATASMANIAQAA 356
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 WVLGVGGTFAGFVLGTFLPVMFMGLHQILTPHIVQMIDETGRLLLPILAMAGAGQVGA 136
QY 357 CLAVFLLAKSK--LKGLAGASVSAVLGTEPAIFGVNLRWPPYIGIGTAAGGALI 414
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
137 ALALWIRCKKXELTEMKALPV-GILGICEPLYGVTLPLGRPFITACIGGGIGGAVI 195
QY 415 ALF-DIKAVAGAGFLGVSDADPMVMFLVCVVTFVIAFGAIAIAYGLYLVR--RNGS 471
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 GMIGNVGAIAIGPSGAALPLISDGKWLGYILGLAAYAGGVFATFFFGIPKDRLEKEEL 255
QY 472 IDPDATAAPVPAGITTKAEAPAEFSDNSTIIQAPLTGEATALSSVSDAMFASCKLGSGV 531
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 ATETATVNEVP-----EKVVPQTSAESVTILSSVADGVTVVSLDKASDPVFAQKMMGEY 309
QY 532 AIVPTKGQLVSPVSGKIYVAFPPSGHAFVRTKAEDGSNVDILMHIGFDTVNLNTHFNPL 591
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
310 FVEPENGHIYSPVTGVWSSVPTKHAIGITTP---NGLEILLHMGINTVLDLGGTPTFDVK 365
QY 592 KKQGEVYAGELLCFEDDAIKAAGYEVTTPVIVSNYKKTGPVNTYGLGEIEAGANLLNV 651
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
366 VTGGOVTPDTLLVAEDLEAIKAKGKTSMTIVLVITNMORVKNFVLEKTKAKAKAPVMDV 425

RESULT 8
US-09-107-532A-6644
; Sequence 6644, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-5007
; INFORMATION FOR SEQ ID NO: 6644:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:

```



```
QY 311 FGLVSPVITGLHQSF-----PPIELEFNQ-----GSFIF 343
| | | | |
DB 237 FGRIKLLIFGLHHIFHAFWFGSKNAAGEIRGDRIFIEIREGAHUTSGKFMQ 296
| | | | |
QY 344 ATASMANIAGAACLAFFLAKSEKLAG-----ASGVSALVIGIITEPAIFGVNLRW 399
| | | | |
DB 297 GEFFVMFGLPAALAIYOTAKENKVVAGLMISAALTSLTIGITEPLEF--SFLFVAP 354
| | | | |
QY 400 FYIGIGTAAIGGALIALFDIKAVAG--AAGFLGVVSD-APDMVFLVCVVTFVIAP 455
| | | | |
DB 355 FLVIRAVLDGLSFLY--LNLVHGVTFSGGFDIVLLGILPNKTAWMLVIPGVIIYAV 413
| | | | |
QY 456 GAIAAYGLYLVRN-----GSIDPA-----TAAPVPAGTTKA----- 488
| | | | |
DB 414 IYFVFRFLIVKENYTPGREDKSVTTTSASQLPFDVLKAMGGRENKHLHACITRLR 473
| | | | |
QY 489 -----EAFAPAEFSN 498
| | | | |
DB 474 IQNEKSKVDVAGLKLGSAGSVLEVNMMQAIIFGPKSDQIKHDMAKIISGEITKPSSETI 533
| | | | |
QY 499 D-----STIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQVSPV 544
| | | | |
DB 534 DEEVSDSHVEDIVETIYAPGHGIIPLSEVPDKVFEKMMGDGIGFVPSDKIVAPF 593
| | | | |
QY 545 SGRIVVAEPGSHAFATRAEDGSNVDILMHIGFDVNLNGTHFNPLKKQGEYKAGELL 604
| | | | |
DB 594 DGVTKIETPKHAIGL-----ESDSGVEVLHIGIDIVKNGEGFESLVNTDEPVTGGQPL 649
| | | | |
QY 605 CEFDIDAKAAGVETVTPVIVSNYK-KTGPVNTYGLGEIEAGANLLNV 651
| | | | |
DB 650 MKTIDLEVLEHAPSIIPTVIITNQEDKT--LTIEDVKQVDPGKAIMTI 695
| | | | |

RESULT 11
US-09-107-532A-5083
; Sequence 5083, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5083:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 656 amino acids
```

```
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...656
SEQUENCE DESCRIPTION: SEQ ID NO: 5083:
US-09-107-532A-5083

Query Match
Best Local Similarity 9.4%; Score 315; DB 4; Length 656;
Best Local Similarity 22.2%; Pred. No. 2.5e-23;
Matches 151; Conservative 96; Mismatches 207; Indels 226; Gaps 27;

QY 118 IPILVGGGLLMAIN-----NVLVAQDL-----FGPQSLVEMFPOIS 153
| | | | |
DB 59 LPLLFAVLGLAGMSVDKSAAALAGLVAFEVFNVLKTDTSVATLLNINIKPESVDPSFAQIS 118
| | | | |
QY 154 GVAEMINLMASAPFAFLPVLVGFATKFRGGNEFLGAGIGMAMVFPFLLVNGYDVAATMTA 213
| | | | |
DB 119 NV-----FIGIAGLIAAALY--NRFEAKLPALSF----- 148
| | | | |
QY 214 GEMPWNSFLGLDVAQAGQGVPLVPLVVSILATIEKFLHRLKMGTAFLTPVTLTLLT 273
| | | | |
DB 149 -----FSGKRLVPILAAFNLLISAVL-----LLVMPFVYSALV 182
| | | | |
QY 274 GFLTEIAGPAMRWYGDLLAHGLOGLYDFGGPVGGLFLGLVSPVITGLHQSPPI--- 330
| | | | |
DB 183 SFAKFL--SLGMWG-----AGLYGF-----FNRLIIP--TGLEQAMNSVFWF 221
| | | | |
QY 331 ELELENQSGSFIFA--TASMANIAG-----AACLAVFFLAKSE-KLKG---LA 373
| | | | |
DB 222 DVAGINDIGKFLASEGTGVTGMVQAGFFIMMFLGPAGAYAIYRNARPENKAKTASLML 281
| | | | |
QY 374 GASGVSAVLGTEPAIFGVNLRWRPFYI--GIGTAAIGGALIALFDIKAVAGLAAG--- 428
| | | | |
DB 282 AAFAFSEFTGTEPLEFSF-MFVAPLVYLHAIET-GISLAVSAFFHWTAGFAFSAGFVD 339
| | | | |
QY 429 -FLGVVSDADPMVNFVCAVVTVI-AFG----- 456
| | | | |
DB 340 YFLSLKNFIANOPLMLIVQGLVTAIVYFVGFFAIFKFLHMTGPREEADLSDDDTATTNS 399
| | | | |
QY 457 ---AAIAYGLV-----VRRNGSIDPD---ATAAP----- 480
| | | | |
DB 400 DNKYAAQASKIYAAALGGADNVTSDNCTTRLRLQVKDTGLIDQNKIKATGVPGMKIIDGK 459
| | | | |
QY 481 ---VPAGT-----TKAEAPAEFSNDSTIIQ-----APLTGEAIAL 514
| | | | |
DB 460 NAOQVIVGTEVQFVADEMAKLHGGAARPAQTNTVKTETFAGETIEQDIAIANGKLIPI 519
| | | | |
QY 515 SSVSDAMPASGKLGSGVAIVPTKGOLVSPVSGKIVVAFPSGHAFATVTKAEDGSNDVILM 574
| | | | |
DB 520 TEVSDDVFAEKMMGDGYAVLPENGEIFSPIAGTTNIPPTKHAVGICQIDA-----GIEVLL 575
| | | | |
QY 575 HIGFDTVNLNGTHFNPLKKQGEYKAGELLCFDDIDAIIKAAGAEVYTPIVVSNYKKTGPV 634
| | | | |
DB 576 HMGINTVDLKGEPTFLYVEEGQKVARGLIALVDLAAIQSAGKNTDMVWVFTNGDKVQSL 635
| | | | |
QY 635 NTYGLGEIEAGANLLNVAKK 654
| | | | |
DB 636 EIEPARDVKNDKIGSVSNK 655
| | | | |
```

## RESULT 12

```
US-09-107-532A-4366
; Sequence 4366, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
```

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESS: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4366:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 585 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...585  
SEQUENCE DESCRIPTION: SEQ ID NO: 4366:  
US-09-107-532A-4366  
Query Match 8.9%; Score 296.5; DB 4; Length 585;  
Best Local Similarity 20.6%; Pred. No. 1.6e-21;  
Matches 136; Conservative 99; Mismatches 189; Indels 237; Gaps 25;  
QY 99 NAWFSEAYKVLADIFVPLIPILVGGLLMANN--VLVAQDLFGPQSLVEMPPQISGVA 156  
DB 40 NGRFVFFVILWSTIVAILATVWPVQAGINNGLWIAE---SKESAPIIAPFLYGLT 96  
QY 157 EMINMASAPFA---FLPVLVGTATKRFGGNEFLGAGIGMAMVFTLVNGYDVAATMTA 213  
DB 97 ERULL---PFGLHMLTIPINT---QLGGTYEILSGA-----QACTQVF 135  
QY 214 GEMPMWSLFGDLVAQAGYQGVLPVLVWSWILATIEKFLHKLMTADFLITPVLTLIT 273  
DB 136 GDPLMLAWATDLV-----NLKGAGDM---SQYEFVLT 165  
QY 274 GFLTFTAIAPMARWGDLLAHGLQGLYDFGGPVGGLLVSPVITGLHOSFPFIELE 333  
DB 166 NWT-----PARKVQOMIGS-----SGILMGLTIA-----MYRNVDPDKS 201  
QY 334 LFNQGSFTATASMANIAQGAACLAFFLAKSEKLGLAGAGSVSAVLGITPAIFGVN 393  
DB 202 KYR---SMVF-----SAALAVF-----LTGVTEPLEF-NF 227  
QY 394 LRLRWPFYIGIGTAAGGALIALFDIKAVAGAGFLGV-----VSIDAP---DMVMFLV 445  
DB 228 MFAVPLIYV--SVTQGAFAWADILPLRVSGNIELLTRPTPLAKAGLIDLLNFVI 285  
QY 446 CAVVTFVIAFGAIAAYGLYLVR-----RNSIDPDATAAPVPAGTT-----486

Db 286 CVILFGVLTFFVA-----NFLIKKFNYATPGRNGYNDSEETPSGSATNADQIIKIHL 341  
QY 487 -----KAEAEAPAEFSND-----STIIQ 504  
Db 342 LGGKENIKVDVACMTRLRVSVEDPGKVGSEEWKVRAGAMGLTVKDKGVQAVYGFKADYJK 401  
QY 505 APL-----TGEAIALSSVSDAMFAS 524  
Db 402 SDIEDLLQSGADIPPTVEATNKNEKNEKHVGLGIEKELFVATGEVIALTDVNDPVFSQ 461  
QY 525 GKLGSVAIVTKGQLVSPVSGKIIVAPPSGHAFAVRTKAEDGSNDVILMHIGFTVNLN 584  
Db 462 KMMGDFAVIPATGVEAAPLSKILSVFPTKEAIGMQT-----AEGAEVLIHMGLDTHMS 517  
QY 585 GTHFNPLKKQGDDEVKAGELLCEFDIDAIAKAAGYEVTTPIVVSNTYKKTGPVNTYIGIETEA 644  
Db 518 QPAFELVSEGOEVVAGTTIAQMNLDIAKNEGKE-TTIIVVFTDDKVNGLIINKLGDTER 576  
QY 645 G 645  
Db 577 G 577

RESULT 13  
US-09-107-532A-5798  
; Sequence 5798, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5798:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 470 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (B) LOCATION 1...470  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5798:  
US-09-107-532A-5798

Query Match 8.1%; Score 271.5; DB 4; Length 470;  
Best Local Similarity 21.1%; Pred. No. 4.3e-19;  
Matches 116; Conservative 78; Mismatches 143; Indels 213; Gaps 21;

QY 208 AATWAGEMPWISFLGLDVAAGYQGVLPVVLVSVWILAEKFLHRLMGADFLITPV 267  
DB 14 AGTVQGDPLWLAWATDLNLKAGAD-----MSQYFVLHNT-----52  
QY 268 LTLITGFLTEAIGPAMRWYGDLLAHGLOGLYDFGGPVGGLFLGLVSPIVITGLHQSF 327  
DB 53 -----PAREKVGQMGIS-----SGILMGMAFA-----MYRNV 79  
QY 328 PPIELFNOGGSFIFATASMANIAOAGACLAFLAKSEKLGKLAGASGVSAYLGTPE 387  
DB 80 DPDKRKYK-----SMYF-----SALAVF-----LTGVTEP 106  
QY 388 AIFGVNLRRLRPFFIGIGTAAIGGALIALFEDIKAVALGAAGFLGV-----VSDA---PD 439  
DB 107 LEF-MEMFAAVPLAVIY--SVIQGAFAFAMADILFLRVESFGNIELLRTPLAKAGLGGD 163  
QY 440 MYMELVCAVTVFAFGAAATAYGL--YLVF-----RNGSIDPDATAAPVP-AGTKA 488  
DB 164 LINEVLC-----VAFG-VVTYFLANFLIKFNATPCRNNGNYDNDSEETVYVNSGTGTA 217  
QY 489 EAE-----VAFG-VVTYFLANFLIKFNATPCRNNGNYDNDSEETVYVNSGTGTA 217  
DB 218 DQVVVQIHLGGKENIKVDYDACHTRLRVSVNDREKVGTEAMKRAGAMGLIVKDNQVQA 277  
QY 495 EFSNDSIIQAPL-----TGEAIALS 515  
DB 278 VYGPADVLKSDIEDLLQSGAEIPEPEMVSTGNGQAGSKOYLGTIEQELVSAASGEVITPLS 337  
QY 516 SVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIIVAPPSPGHAFVATKAEAGSNVDILMH 575  
DB 338 EVKDPVFSQKMWGDFAVITSRVAPVPIAGKVTISFPSKHAIGLETK--DG--IEVLII 393  
QY 576 IGFTVNLNGTHFNPLKKQGEVKAAGELLCFDDIDAIAKAAGYEVTTPIVVSNNKKTGPVN 635  
DB 394 MGIDTVQMNOPAFETIAVKEGQEVGTGKLAEMDLVDIQNEEKDFTIMIVFTN-DKVEEV 452  
QY 636 TYGLGEAEAG 645  
DB 453 IKQLGTWTAG 462

RESULT 14  
US-09-134-001C-3134  
; Sequence 3134, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3134  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3134

Query Match 8.0%; Score 267; DB 4; Length 170;  
Best Local Similarity 39.8%; Pred. No. 2.2e-19;  
Matches 53; Conservative 31; Mismatches 45; Indels 4; Gaps 2;

QY 503 IQAPLTGTEAIALSVSDAMFASGKLGSGVAIVPTKGOLVSPVSGKIIVAPPSPGHAFVAT 562

DB 22 IYAPLTGEVYKIEDIDPVPFAQKMMGEGGINTEGEVVSPIEGKVDNVFPTKH--AVGL 79  
QY 563 KADGNGNVDILMHIGFTDVTNLTGTHFNPLKKQGEVKAAGELLCFDDIDAIAKAAGYEVTT 622  
DB 80 KAENG--LELLVHIGLDTVQLDQKGFELVESGDDIKIGDPLIRFDLEYINNNAKSIISP 137  
QY 623 IVVSNNKKTGPVN 635  
DB 138 IITNSDQTESIH 150

RESULT 15  
US-09-252-991A-18115  
; Sequence 18115, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18115  
; LENGTH: 686  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18115

Query Match 6.9%; Score 230; DB 4; Length 686;  
Best Local Similarity 32.2%; Pred. No. 1.6e-14;  
Matches 59; Conservative 40; Mismatches 68; Indels 16; Gaps 6;

QY 488 AEAAPAE-----FNSNSTIIQAPLTGTEAIALSVSDAMFASGKLGSGVAIVPTKGOLVS 542  
DB 1 AKSPEPTEPERAPMNNKNLAKPLSGVPVPLNRVPDPVFSSGTGEGIALDPLNDCLHA 60  
QY 543 PVSGKIIVAPPSPGHAFVATKAEAGSNVDILMHIGFTDVTNLTGTHFNPLKKQGEVKAAG 602  
DB 61 PCAGLVSHLARTRHLSLR--ADNGA--ELLHVGLDVTVOLQGEFGEALVEEGARVIEGQ 116  
QY 603 LLCEEDIDAIAKAAGYEVTTPIVVSNNK--KTGPNVTYGLGEIEAGANLLNVA--KKEAVP 658  
DB 117 PLLRFDLDRVARGSLITVMILTNGDGFQVRPLTT---NPVEVGAPLQLUSPEAKAQR 173  
QY 659 ATP 661  
DB 174 ANP 176

Search completed: October 1, 2003, 05:05:41  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 05:04:06 ; Search time 217 Seconds  
(without alignments)  
481.930 Million cell updates/sec

Title: US-10-019-284A-2  
Perfect score: 3342  
Sequence: 1 MDHKLQAQRILRDIGGEDNI.....IEAGANLLNVAKKEAVPATP 661

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues 587654

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3310	99.0	661	10 US-09-738-626-6404	Sequence 6404, Ap
2	956	28.6	627	9 US-09-815-242-13467	Sequence 13467, A
3	725.5	21.7	683	10 US-09-738-626-6961	Sequence 6961, Ap
4	550	16.5	484	9 US-09-815-242-10809	Sequence 10809, A
5	538	16.1	455	9 US-09-815-242-13794	Sequence 13794, A
6	522	15.6	484	9 US-09-815-242-12272	Sequence 12272, A
7	490.5	14.7	439	9 US-09-815-242-5660	Sequence 5660, Ap
8	472	14.1	545	15 US-10-156-761-11459	Sequence 11459, A
9	438	13.1	474	9 US-09-815-242-10243	Sequence 10243, A
10	368.5	11.0	679	9 US-09-815-242-5658	Sequence 5658, Ap
11	368.5	11.0	681	9 US-09-815-242-12270	Sequence 12270, A
12	321	9.6	719	9 US-09-815-242-12942	Sequence 12942, A
13	319	9.5	687	9 US-09-815-242-5822	Sequence 5822, Ap
14	308.5	9.2	648	9 US-09-741-669-330	Sequence 330, App
15	308.5	9.2	648	9 US-09-815-242-10089	Sequence 10089, A

16	290.5	8.7	551	9 US-09-815-242-10500	Sequence 10500, A
17	233.5	7.0	640	9 US-09-815-242-10786	Sequence 10786, A
18	232	6.9	280	9 US-09-815-242-12479	Sequence 12479, A
19	226.5	6.8	842	9 US-09-815-242-11950	Sequence 11950, A
20	221.5	6.6	263	9 US-09-815-242-13018	Sequence 13018, A
21	221.5	6.6	263	9 US-09-815-242-13152	Sequence 13152, A
22	220.5	6.6	135	9 US-09-815-242-5842	Sequence 5842, Ap
23	209	6.3	149	15 US-10-156-761-14503	Sequence 14503, A
24	201.5	6.0	142	9 US-09-815-242-5718	Sequence 5718, Ap
25	187	5.6	657	9 US-09-815-242-12301	Sequence 12301, A
26	183.5	5.5	98	9 US-09-939-980-419	Sequence 419, App
27	182	5.4	643	9 US-09-815-242-5394	Sequence 5394, Ap
28	173.5	5.2	650	9 US-09-815-242-13341	Sequence 13341, A
29	173.5	5.2	650	11 US-09-769-787-95	Sequence 95, Appl
30	173	5.2	704	15 US-10-156-761-11225	Sequence 11225, A
31	172.5	5.2	658	9 US-09-912-020-352	Sequence 352, App
32	168.5	5.0	422	15 US-10-156-761-12701	Sequence 12701, A
33	157	5.0	639	12 US-10-238-075-607	Sequence 607, App
34	160.5	4.8	448	11 US-09-769-787-56	Sequence 56, Appl
35	157.5	4.7	482	9 US-09-815-242-5401	Sequence 5401, Ap
36	157.5	4.7	488	9 US-09-815-242-12564	Sequence 12564, A
37	156	4.7	585	9 US-09-815-242-11923	Sequence 11923, A
38	150.5	4.5	556	9 US-09-815-242-11023	Sequence 11023, A
39	141.5	4.2	454	9 US-09-815-242-10731	Sequence 10731, A
40	140.5	4.2	510	9 US-09-815-242-13474	Sequence 13474, A
41	140	4.2	534	15 US-10-156-761-12074	Sequence 12074, A
42	138.5	4.1	505	11 US-09-769-787-120	Sequence 120, App
43	138.5	4.1	510	9 US-09-815-242-13607	Sequence 13607, A
44	138	4.1	475	9 US-09-815-242-13774	Sequence 13774, A
45	138	4.1	502	15 US-10-156-761-8917	Sequence 8917, Ap

## ALIGNMENTS

## RESULT 1

US-09-738-626-6404  
; Sequence 6404, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIALI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/250988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6404  
; LENGTH: 661  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6404

Query Match 99.0%; Score 3310; DB 10; Length 661;  
Best Local Similarity 98.9%; Pred. No. 1e-290;  
Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDHKLQAQRILRDIGGEDNIIVAAAHCAATRLRLVLTCKDVRQSLDDDPDLKGTFTGGM 60

```
Db 1 MDKDLAQRLDRIGGEDNIVAAAHCAATRLRLVLKDTKDVRQSLDDDDPKLGTFTETGGM 60
QY 61 FOIIVPGDGVHVFKELDATSKDIADVSTEQKDVVANNANWFSRAVKVLADIFVPLPI 120
Db 61 FOIIVPGDGVHVFKELDATSKDIADVSTEQKDVVANNANWFSRAVKVLADIFVPLPI 120
QY 121 LVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATK 180
Db 121 LVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATK 180
QY 181 REGGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLPVLV 240
Db 181 REGGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLPVLV 240
QY 241 VSNILATIEKFLHKLRLMGTADELITPVLTLTLTGTFTFAIGPAMRWVGDLLAHGLQGLY 300
Db 241 VSNILATIEKFLHKLRLMGTADELITPVLTLTLTGTFTFAIGPAMRWVGDLLAHGLQGLY 300
QY 301 DFGGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGAACLAY 360
Db 301 DFGGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGAACLAY 360
QY 361 FFLAKSEKIKLAGASGSAVIGITEPAIFGVNLRNRPFFYIGTAAIGGALIALFDIK 420
Db 361 FFLAKSEKIKLAGASGSAVIGITEPAIFGVNLRNRPFFYIGTAAIGGALIALFDIK 420
QY 421 AVALGAAGFLGVVYSDAPDMVFLCAVVFTEAFGAALAYGLYLVRNGSIDPDATAP 480
Db 421 AVALGAAGFLGVVYSDAPDMVFLCAVVFTEAFGAALAYGLYLVRNGSIDPDATAP 480
QY 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGKLSGVAIVPTKQGL 540
Db 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGKLSGVAIVPTKQGL 540
QY 541 VSPVSGKIVVAPPSGHAFVRKAEDGSNVDILMHIGDPTVNLNGTHFNPLKKQGDVKA 600
Db 541 VSPVSGKIVVAPPSGHAFVRKAEDGSNVDILMHIGDPTVNLNGTHFNPLKKQGDVKA 600
QY 601 GELLCEFDIDAIAKAAGYEVTTPIVSNYKKTGPVNTYGLGETEAGANLLNVAKKEAVPAT 660
Db 601 GELLCEFDIDAIAKAAGYEVTTPIVSNYKKTGPVNTYGLGETEAGANLLNVAKKEAVPAT 660
QY 661 P 661
Db 661 P 661
```

## RESULT 2

US-09-815-242-13467

Sequence 13467, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

```
QY 1 MDKDLAQRLDRIGGEDNIVAAAHCAATRLRLVLKDTKDVRQSLDDDDPKLGTFTETGGM 60
Db 1 MDKDLAQRLDRIGGEDNIVAAAHCAATRLRLVLKDTKDVRQSLDDDDPKLGTFTETGGM 60
QY 61 FOIIVPGDGVHVFKELDATSKDIADVSTEQKDVVANNANWFSRAVKVLADIFVPL 117
Db 61 FOIIVPGDGVHVFKELDATSKDIADVSTEQKDVVANNANWFSRAVKVLADIFVPL 115
QY 118 IPILVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGT 177
Db 118 IPILVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGT 165
QY 178 ATKREGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLP 237
Db 178 ATKREGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLP 221
QY 238 VLVYVSNILATIEKFLHKLRLMGTADELITPVLTLTLTGTFTFAIGPAMRWVGDLLAHGLQ 297
Db 238 VLVYVSNILATIEKFLHKLRLMGTADELITPVLTLTLTGTFTFAIGPAMRWVGDLLAHGLQ 281
QY 298 GLYDFGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGA 356
Db 298 GLYDFGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANIAQGA 341
QY 357 CLAVFFFLAKSEKIKLAGASGSAVIGITEPAIFGVNLRNRPFFYIGTAAIGGALIAL 416
Db 357 CLAVFFFLAKSEKIKLAGASGSAVIGITEPAIFGVNLRNRPFFYIGTAAIGGALIAL 401
QY 417 FDIKAVLGAAGFLGVVYSDAPDMV-----MFLCAVVFTEAFGAALAYGLYLVR 468
Db 417 FDIKAVLGAAGFLGVVYSDAPDMV-----MFLCAVVFTEAFGAALAYGLYLVR 449
QY 469 NGSIDPDATAAPVPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGKL 527
Db 469 NGSIDPDATAAPVPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGKL 505
QY 528 GSGVAIVPTKQGLVSPVSGKIVVAPPSGHAFVRKAEDGSNVDILMHIGDPTVNLNGTH 587
Db 528 GSGVAIVPTKQGLVSPVSGKIVVAPPSGHAFVRKAEDGSNVDILMHIGDPTVNLNGTH 561
QY 588 ENPLKKQGDVKAAGYEVTTPIVSNYKKTGPVNTYGLGETEAGANLLNVAKKEAV 644
Db 588 ENPLKKQGDVKAAGYEVTTPIVSNYKKTGPVNTYGLGETEAGANLLNVAKKEAV 618
QY 645 GANLNV 651
Db 645 GANLNV 625
```

US-09-815-242-13467

Query Match 28.6%, Score 956; DB 9; Length 627;

Best Local Similarity 33.3%; Pred. No. 1.2e-77;

Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;

## RESULT 3

US-09-738-626-6961

Sequence 6961, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

```

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6961
; LENGTH: 683
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6961

Query Match          21.7%; Score 725.5; DB 10; Length 683;
Best Local Similarity 30.9%; Pred. No. 9.9e-57;
Matches 216; Conservative 119; Mismatches 286; Indels 79; Gaps 18;

QY 7 AQRILRDIGGDNVAAAHACATRLRLVLTQKDVDRQSLDDDDPDLKGTFTGTG--GMFQII 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 SQHILENLOGGPDNTSMTHCATRLRFQVQDSIVDQGEIDSPSVLGVVPGSGTGM-QVV 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 VPGDGDVHFREL-----DDATSKDIANVSTOLKDVYANNANWFSRAVKVLADI 113
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 MG-GSVANYQOEILKDKMKHFDGEATES----SSKXGYGVGRKYSWIDYAEFEESDT 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 114 FVPLIPVLVGGGLMAINNVLVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVL 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 FRPILWALLGASLIITL---LVADTFGLQDFRAPDEQPTTYVFLHSMWRSVYFLPIM 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 174 VGFTATKRFGEFNEFLGAGIGAMVFPTLVNVDYVAANTAGEMPWMSLFGLDVAQAQYQG 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 VGATAARGLGANEWGAIPAALLTP-----EFLAGSAGDTV--TVFGLPMVLYNDYSG 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 234 TVLPVLVYSWILATIEKFLHRLMGTADFLITPVLITLLTGFLFAIGPAMRWGDLA 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 QVFPPLIAAGLYWVEKGLKIIPEAVQMVFPFESLLIMTPATAFLLGPGIGVNGIS 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 294 HGLQGLYDFGPGVGGLLFLGLYSPITVITGLHOSPPPIELELFNO--GGSFIFATASMANIA 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 NLLEAINNFPSPILSIVIPILYPLVPLGLHPLNATMIQNTILGYDFIQGPMGAWNFA 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 353 QGACLAFFFLAKSKPLAGLACAS--GYSAVL--GITEPAIFGVNLRWRPFYIGTAA 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 CEGLVTVGFLLSIKERNAMQVSLGGMLAGLLGISEPSLYGLLRPKTYIFRLLPGCL 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 409 IGGALIALFDIKAVAGALGFLGVVSDAPDMVFLVCVVTFVIAF--GAIAIAGLYLVR 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 AGGIVMGIFDIKAYAFVFTSLTIPAMD--PWLGYTIGIAVAFVYSMFLVALDY-----R 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 468 RNSGID-----PDAT-AAPV-----PAGTTKAEAEAPA 494
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 SNEERDEARAKVAADQAEEDLKAEANATPAAPVAAGAGAGAGAGAGAATAVAAPK 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 495 EFSNDSITIQAPLTGEAIALSSVDSAMFASGKLGSGVAIVPTKGLVSPVSGKIVAPPS 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 LAAGEWVDIVSPLECKAIPLSVPDPPIFAAGKLGPGIAIQPTGNTVAVAPADATVLYQKS 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 GHAFVATKAEDGSDVNDILMHGFDFTVNLNGTHFNPLKKQGDVFKAGELLCEFDIDAICA 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

; APPLICANT: HASSELBECK, ROBERT
; APPLICANT: OHLESEN, KARI L.
; APPLICANT: ZYSKIND, JUDITH W.
; APPLICANT: WALL, DANIEL
; APPLICANT: TRAWICK, JOHN D.
; APPLICANT: CARR, GRANT J.
; APPLICANT: YAMAMOTO, ROBERT T.
; APPLICANT: XU, H. HOWARD
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10809

Query Match          16.5%; Score 550; DB 9; Length 484;
Best Local Similarity 29.7%; Pred. No. 4.5e-41;
Matches 149; Conservative 99; Mismatches 180; Indels 74; Gaps 16;

QY 4 KDLAQRILRDIGGEDNIVAAAHACATRLRLVLTQKDVDRQSLDDDDPDLKGTFTGTGFMQI 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 KRIADGTYKEVGQENVDKVIHCHMTRVRMDIRDYDKVIEGLKKIDGVVMGVWEDDTL-QV 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 64 IVPGDGDVHFRELDD-----ATPSKDIAVSTEQLDKW-----A 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 VVPGGVTKVQAQWVDQAGVKLGEFPFHGTTD--ASAGKSGKDLVEEKAQKQAKQ 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 98 NNANWFSAVKVLADIFVPLIPILVG---GGLLMANNVLVAQDLFGP--QSLVEMFPQI 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 NNTSPKKVLKATISSIFVPMIPAFVAGIIGIAGVAVMSNLVWAGDISASWQYI----- 178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 SGVAEMINLMASAPFAFLPVLVGTFTATKRFEGGNEFLGAGIGAMVFPPTLVNVDVAATMT 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 ----DVUNIKNGIFAYLALYTGINSVFCATPAAGVIG-----AVTML 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 AG---EMPWMSLFGDLVAQYQGTVLPVLVWVSNILATIEKFLHRLMGTADFLITPVL 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 221 TGMNPDAPISNIFTGTTLSAG--QGGIIGVIFAVALLSLEKQLHKIVPESIDIIVTPTIS 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 270 LLLTGLTFTFAIGPAMRWGDLAHGLQGLYDF-----GGPVGGLLGLVYSPIVITGLHQ 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 280 LLVIGLATIELIMP-----VAGASINGLVIGINVLEKGMVAGTGLTGLTLPWVMEGLHQ 335
QY 326 SFPIELELENOG-SPIFATASANTAOAGACLAFFELAKSEK--LKGLAGASGYSAVL 382
Db 336 ILPIFIHEMINQGTMLLILFALWAGAGQVGAALWIRCKSKKLIVEMIKGALPV-GIL 394
QY 383 GITEPAIFGVNLRWPFYIGIGTAAIGGALIALF-DIKAVALGAAGFLGVVSIDAPDMV 441
Db 395 GIGEPYIVGTLPGRPFITACIGGGIGGAGVAFAGFNGVGAIGAIPSGVALIPLIANQWL 454
QY 442 MFLVCAVTVFIAPGAIAAYGL 463
Db 455 AYVLGLLAAYAGGFVATLFFGI 476

RESULT 5
US-09-815-242-13794
; Sequence 13794, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13794
; LENGTH: 455
; TYPE: PRN
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(455)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13794

Query Match 16.1%; Score 538; DB 9; Length 455;
Best Local Similarity 30.7%; Pred. No. 5e-40;
Matches 141; Conservative 91; Mismatches 192; Indels 36; Gaps 10;
```

```
QY 6 LAQRIIDIGDENIVAAAHCAHCAHRLRLVLDKTDVDRQSLDDDDPDLKGTETGGMFOIIV 65
Db 9 LASDILRIGIGGEQILLENLCMTVRVEVQDDSLDIPRLKALPGVSYVKQGEHQHOLIV 68
QY 66 GPGVDVHVKELDDATSKDIAVSTEQCLKDVYVANNANW-----FSRAVKVLADIFVPL 117
Db 69 GPGKAAQV-----DAMRVQTRAGVKFDDAMARTKSEAKAKYAPNSDALUKLANFVPL 124
QY 118 IPIVLGGGLLMAINNVLAQDLFGPQSLVEMFFPOISGVAEMINLMASAPFAFLVVLVGT 177
```

```
Db 125 IFAFIASGLITGIINILKRPDIVG--DVAVHYPNLLG---LMGIFGSVAFSIMNVLGVN 179
QY 178 ATKRRGGNEFLCAGTGMAMVFTLVNGVDVAATMTAGEMPWMSLFGDLVQAQGYOCTVLP 237
Db 180 TAKVFGGSQALG-GV-----MAGILSSPOLAQITLFG-EALOPG-RGGVIA 222
QY 238 VLVVSWILATIEKFLHKLRLMGADFLITPVLTLTLLTFTIAIGPAMRWGDLLEHGIQ 297
Db 223 VLLVVALMCWIERQFRKLLPGSLELLNPLLTVTITGAIVAVALQPLGWNISDAHAAS 282
QY 298 GLYDFGPPVGGLLFGLVYSPVITGHOSFPFIELEFN-QGGSFIFATASVANTAQA 356
Db 283 WAIDRGGFLVGAVLAGTFLPLVLIGLHQLVHVELVQAHXYNALFFILAMAGVQIGA 342
QY 357 CLAVFFLAKSEKLGLAGASGVSALVIGITEPAIFGVNLRWPFYIGITGAIGALIAL 416
Db 343 AIAVLMKTRNARLKKVIRGALPVGLLIGICEPLIFGVTLPLGKPFPGACLGAVGALLISY 402
QY 417 FDIKAV-ALGAAGFLGVVSIDAPDMVFLVCAVTVTFIAF 455
Db 403 WKVATVITFGISGLPLALTIVAGKVLFLVLLGYLIAIAGF 442

RESULT 6
US-09-815-242-12272
; Sequence 12272, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12272
; LENGTH: 484
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-12272

Query Match 15.6%; Score 522; DB 9; Length 484;
Best Local Similarity 29.3%; Pred. No. 1.6e-38;
Matches 155; Conservative 84; Mismatches 188; Indels 102; Gaps 17;
```

```
QY 4 KDLAQRIIDIGDENIVAAAHCAHCAHRLRLVLDKTDVDRQSLDDDDPDLKGTETGGMFOI 63
Db 5 QQLAERLIIAAGVGMNIDSVNMCMTVRIRKLVJDNKNVDQDLRHIDGVNGVYH-DERIQV 63
QY 64 IVGPGVDVHVKELDDATSKDIAVSTEQCLKDVYVANNANW----- 101
```

Db 64 VVPGTGVKVNANMAE-----LSGVKLGDTIPHNHNDSEKMDYKSYAADKAKANKEAH 116  
QY 102 -----WFSRAVKVLADIIVPLIPLVVG-----GGLMAINNVVAQDLFGPSLVEMFP 150  
Db 117 KAKQKNGKLNKVLKSANIFILPIPAFIGAGLIGGIAAVLSNLMVAGYISG-----A 168  
QY 151 QISGVAEMINLMASAPFAFVLVGTATKRFEGNEFLGAGIGMAMVFPPLVNGYDVAAT 210  
Db 169 WITQLITVFNVIKDMGLAYLAIFTGINAAKEFGATPGLGGVIGGT-----TLLTG----- 218  
QY 211 MTAGEMPWMSLFGDLVAQAGYOGTVPLVAVSVLTAIEKFLHKLMTGADFLITPVLTL 270  
Db 219 -IAGKNILMNVFTGEPLQPG-QGGIIGVIFAVWILSVLRKLNKIPVNAIDIIVTPTIAL 276  
QY 271 LLTGFLTFIATGPAWRWGDLLAHGLQGLYDFGPGVGLLFLGVYSPVITVGLHQSPFPI 330  
Db 277 LIVGLLTIIFMPLAGFVSDLSVSVNGIISIGVFSFGFIIGASFLPLVMLGLHHIFTPI 336  
QY 331 ELEFNQ--GGSFIFATASMANIAQAACLAFFLAKSEKL--KGLAGASGVSAVLGITPE 387  
Db 337 HIEMINGSGATYLLPIAAMAGAGQVGAALALVWRCKRNTTLRNTLKGALPV--GFLGIGEP 395  
QY 388 AIFGVNRLRPWFIY-----GIGTAATGGALIALFDIKAVAGAGFLGVVSDADPMVM 442  
Db 396 LIYGVTLPLGRPPLTACIGGIGGAVIGG-----IGHIGAKAIGPSG-VSLPLISDNMYL 450  
QY 443 FLVCANVTFFIAGAAATAYGL---YLVRNGSIDPDNATAAPVAGTTKA 488  
Db 451 GYIAGLL-----AAYAGGVCTYLF-----GTTKA 475

## RESULT 7

US-09-815-242-5660  
; Sequence 5660, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5660  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5660

Query Match

14.7%; Score 490.5; DB 9; Length 439;

Best Local Similarity 30.3%; Pred. No. 9.5e-36;  
Matches 131; Conservative 71; Mismatches 166; Indels 65; Gaps 11;  
QY 4 KDLAQRIIRDIGGDNIVAAAHACATRLRLVLDKTDVDRQSLDDDDDLKGTETGGMFQI 63  
Db 5 QQLAEITIAAAGVGNIDSVNMCNTRVRKIVLDENKNVDQDELHIDGVMGVIH-DERIQV 63  
QY 64 IYVPGDVDHVFRELDODATSKDIAVSTEQDKDVYANNAN----- 101  
Db 64 VVPGTGVKVNANMAE-----LSGVKLGDTIPHNHNDSEKMDYKSYAADKAKANKEAH 116  
QY 102 -----WFSRAVKVLADIIVPLIPLVVG-----GGLMAINNVVAQDLFGPSLVEMFP 150  
Db 117 KAKQKNGKLNKVLKSANIFILPIPAFIGAGLIGGIAAVLSNLMVAGYISG-----A 168  
QY 151 QISGVAEMINLMASAPFAFVLVGTATKRFEGNEFLGAGIGMAMVFPPLVNGYDVAAT 210  
Db 169 WITQLITVFNVIKDMGLAYLAIFTGINAAKEFGATPGLGGVIGGT-----TLLTG----- 218  
QY 211 MTAGEMPWMSLFGDLVAQAGYOGTVPLVAVSVLTAIEKFLHKLMTGADFLITPVLTL 270  
Db 219 -IAGKNILMNVFTGEPLQPG-QGGIIGVIFAVWILSVLRKLNKIPVNAIDIIVTPTIAL 276  
QY 271 LLTGFLTFIATGPAWRWGDLLAHGLQGLYDFGPGVGLLFLGVYSPVITVGLHQSPFPI 330  
Db 277 LIVGLLTIIFMPLAGFVSDLSVSVNGIISIGVFSFGFIIGASFLPLVMLGLHHIFTPI 336  
QY 331 ELEFNQ--GGSFIFATASMANIAQAACLAFFLAKSEKL--KGLAGASGVSAVLGITPE 387  
Db 337 HIEMINGSGATYLLPIAAMAGAGQVGAALALVWRCKRNTTLRNTLKGALPV--GFLGIGEP 395  
QY 388 AIFGVNRLRPWFIY-----GIGTAATGGALIALFDIKAVAGAGFLGVVSDADPMVM 442  
Db 396 LIYGVTLPLGRPPLTACIGGIGGAVIGG-----IGHIGAKAIGPSG-VSLPLISDNMYL 450  
RESULT 8  
US-10-156-761-11459  
; Sequence 11459, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 11459  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-11459

Query Match

14.1%; Score 472; DB 15; Length 545;  
Best Local Similarity 25.9%; Pred. No. 6.3e-34;  
Matches 152; Conservative 79; Mismatches 223; Indels 132; Gaps 14;

QY 3 HKDLAQRIIRDIGGDNIVAAAHACATRLRLVLDKTDVDRQSLDDDDDLKGTETGGMFQ 62  
Db 6 HSATAAAILPLYGGAANTVSAHCWTRLRLGLADSLYDEESLRALPAVLGVWD-DDTYQ 64  
QY 63 IIVPGDGDHVFRELDODATSKDIAVSTEQDKDVYANNAN----- 97

Db 65 IVLGGIVAKVTPPEFKLLASSETTEAPANALPVYTAGAPTFAAPTAAEGPATSETPVAEP 124  
QY 98 -----NNANWFSRAYKVLADIFVPLIPILVGGGLMAI 130  
Db 125 PARGSAFCTEAARLAARGAALKAAQORHATPVRLRLRIANIFVPLIPALIGCVIAG 184  
QY 131 NNVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFTATKRFEGNEPLGA 190  
Db 185 NGLLNLGW-----LPSLTALAAV-----ASGFMALIAVFGVHTAKEFGTPILOG 232  
QY 191 GIGMAVFPVLVNGYDVAATMTAGEMPWMSLFGDLVDAQYQGTVLPVLVSVILATIEK 250  
Db 233 AVAAVIVYAGVAKVTFVGTTLSPG-----QGVLCALAAALLGTVVEK 275  
QY 251 FLHRLMGTADFLTPVLTLLTGLTFLFI-----AIGPANWVWGDLLAHGLOGL 299  
Db 276 WCRTWVPTDLVLTPTLLTGLTGLVLYGLMYAAGAATSAIGTAATW-----L 324  
QY 300 YDFGGPVGGLFGLVYSPVITVGLHQSFPPFIELEFNQGG-SFIFATASMANIAQAACL 358  
Db 325 LSTGTGVFAGLTGLFLPLVMLGLHQAALPLHTLIEQOQYTVLLPVLAMAGAGQVGAAL 384  
QY 359 AVF-LAKSEKLKLAGASGVSALVIGITEPAIGVNLRLRWPYVIGIGTAIGGALIALF 417  
Db 385 AVVRLRHDSLRITISALPAGLGVGEPLTVGSPLGRPFLTACAGGAAGAFVGLF 444  
QY 418 D-----IKAVALGAAGFL-----GVVSDAPDMVMFLVCAVTFVIAFGAAIAYGL 463  
Db 445 SMLGDRIGSTAIGSPGWALPPLLAGSGLL-----PSAAIYAGGLITGYAVGFCATYFFGS 500  
QY 464 YLV-----RRNSIDPDATAAPVPACTTKAEAEAPAFNSDSTI 502  
Db 501 FGUVGGVCEAERTDIAAPTCTAAPTMT-TTPTDTAAPHDPDEGNRTL 545

RESULT 9

US-09-815-242-10243  
; Sequence 10243, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10243  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Escherichia coli

US-09-815-242-10243  
Query Match 13.1%; Score 438; DB 9; Length 474;  
Best Local Similarity 27.8%; Pred. No. 6e-31;  
Matches 139; Conservative 84; Mismatches 213; Indels 64; Gaps 17;  
QY 5 DIAQRLTRDIGEDNIVAAHCASTRRLRLVLTDFKQVDRQSLDDDDPDLLKGTETGGMFOII 64  
Db 8 ELNLTILTRVGGPFIASCNCMTLRUGVHOSLVD-PNITKLGKGVILTSQVQV 66  
QY 65 VGPQGDYDVHFEK-----LDDATSKDIA-VSTEOLKDVANNANWFSRAVKVLADIFVPLIP 119  
Db 67 FPGKAHRAAKAMSELLGEAPVQDAEIAAQNKRKAKAKOTSGVQOFLAKPATIIFTPLIP 126  
QY 120 ILVGGGLLMAINNVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFTAT 179  
Db 127 GRTAAGLLGI-ATLIATVNVHPADAQGTLPD---ALNFMKVFSGKLTFLVLVGYNAA 182  
QY 180 KRFGGNEFLGAGIMAVFPTLVNGYDVAATMTAGEMPWMSLFGDLVDAQYQGTVLPVL 239  
Db 183 QAFG-----GTGVNGAIIAALFLLGYNPAAT-TGYAGFHDFFGLPIDP---RGNIIIGVL 233  
QY 240 VVSWILATIEKFLHRLMGTADFLTPVLTLLTGLTFLFIAIGPAMRWGDLAHGLOGL 299  
Db 234 IAAWACARIEGMVRRFMPDLDMLTSLTLLTATLAYLIIMPLGGW-----LFEQMSWL 289  
QY 300 YDF--GGPYGGLFGLVYSPVITVGLHQSFPPFIELELF-NOGGSFIFATASMANIAQA 356  
Db 290 FMHLNSNPFCAVLAGLFLIAVFGVHQGFIPVYLAALMSQGFNSLFTPLSNAGAGQVGA 349  
QY 357 CIAVFFLAKSEK-LKLAGASGVSALVIGITEPAIGVNLRLRWPYVIGIGTAIGGALIA 415  
Db 350 ALALYWRAPHSALRSQVRGAIIPGLLVGEPLIYGVTLPRMKPFV---TACLGGAAGG 405  
QY 416 LF-----DIAKVAL--GAAGFLGVVSDAPDMVMFLVCAVTFVIAFGAAIAYGL 453  
Db 406 LFGLIAWGLPMGLNSAFPSGLVALPLMTSAQGLIPAMAVYAGIILVAVVCGFI-FTT 464  
QY 454 AFGAAIAYGLYLVRNGSID 473  
Db 465 LFQC-----RNVNLD 474  
RESULT 10  
US-09-815-242-5658  
; Sequence 5658, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10243  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Escherichia coli

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5638  
; LENGTH: 679  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-3658

Query Match 11.0%; Score 368.5; DB 9; Length 679;  
Best Local Similarity 23.3%; Pred. No. 2.1e-24;  
Matches 160; Conservative 85; Mismatches 241; Indels 201; Gaps 23;  
QY 103 FSRVAVLADIFVPLIPILVGGGLMAINNVLAQDL--FGQSLVEMFPOISGVAEMIN 160  
Db 4 FGQQRIGKALMLP--VAILPAAGLLAIGTAIOGEALQHYLP-----FIQNGGVQNVAK 56  
QY 161 LMASAP---FAFLPVLVGTATKRFGGNEFLGAGIGMAMVPTLVNGVDVAA----- 209  
Db 57 LMTAAGSIIFENLPMIFA-----LGVAILGA-----GGDGVAAIAAFVGYI 97  
QY 210 --TWTAGEM-----PMSLFGLDVAAGAGYQGTVPVLVSVWILATIEFLHK 254  
Db 98 INKMTGDFLOVTPKNVTPDASGYASILGIPTLQTVFGGIIIGALAAW---CYNKFYNI 154  
QY 255 RLMTGADFL---ITPVLTLTLTGFLTAIGPAMRWVGDLLAHGLQ---GLYDFGGPV 306  
Db 155 NLPSYLGFFAGKRFVPIV-MATTSFI--LAPPMALIW--PTIOTGLNATSTGLDSDNTGV 209  
QY 307 GGLLGLVYSPIVITGLHQSF-----PPIELFNOG-----G 339  
Db 210 AVFLFGFIKRLIPFGLHHIFHAPFWFEGSKNAAGEIIHGDDRIETIEQIREGAHLTAG 269  
QY 340 SFIFATASMANIAQGAACLAFFFLAKSEKLGAGAGVSA---VLGITEPAIFGVNLR 395  
Db 270 KMQGEFFVMMFGLPAAALAIYHSKPKENKVVAGLMSAALTSLTGTITPLEFSFLV 329  
QY 396 LWPFFYIGTAAIGGALIALFDIKAVAGLAAAGFLGVVSIDA--PDMVMF-----LVCA 447  
Db 330 APLLFFIHAVLDGLSFLTLVLDVHLGYTFSGGFDYVLLGVLPNKQTQWMLVIPVGLVYA 389  
QY 448 VVTF-----VIAFGAAIAYGLY-----LVR- 467  
Db 390 VIYFVFRFLIVKLKYKTPGREDKQSQAVTASATELPYAVLEAMGGRANIKHLDACITRL 449  
QY 468 -----RNGSIDPDATAA 479  
Db 450 RVEVNDKSKVDVPGLKDLGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGQVVENPTM 509  
QY 480 PVPACTTKAEAPAEAFESNDSTIIQAPLTGEAIALSSVSDAMPASGKLGSGVAIVPTKGQ 539  
Db 510 EDDKDETIVVAEDKSAISELSHVHAPLTGEVTPLEVPDQVFEKMMGMDGIAIKPSQGE 569  
QY 540 LVSPYSGKIVAFPSGHAFAYRKAEDGNSVDILMHIGFDTVNLGNHFNPLKKGQGEVK 599  
Db 570 VRAPENGKVMQIFPKHAIGL---VSDSGLELLIHIGLDIVKLNGBEFTLHVEEGQEVK 625  
QY 600 AGELLCEFDIDAIRKAAGVEVTPPIVVS 626  
Db 626 QGDLINFDLDYIRNHAKSIDTITPIIVT 652

## RESULT 11

US-09-815-242-12270  
; Sequence 12270, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12270  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12270

Query Match 11.0%; Score 368.5; DB 9; Length 681;  
Best Local Similarity 23.3%; Pred. No. 2.1e-24;  
Matches 160; Conservative 85; Mismatches 241; Indels 201; Gaps 23;  
QY 103 FSRVAVLADIFVPLIPILVGGGLMAINNVLAQDL--FGQSLVEMFPOISGVAEMIN 160  
Db 6 FGQQRIGKALMLP--VAILPAAGLLAIGTAIOGEALQHYLP-----FIQNGGVQNVAK 58  
QY 161 LMASAP---FAFLPVLVGTATKRFGGNEFLGAGIGMAMVPTLVNGVDVAA----- 209  
Db 59 LMTAAGSIIFENLPMIFA-----LGVAILGA-----GGDGVAAIAAFVGYI 99  
QY 210 --TWTAGEM-----PMSLFGLDVAAGAGYQGTVPVLVSVWILATIEFLHK 254  
Db 100 INKMTGDFLOVTPKNVTPDASGYASILGIPTLQTVFGGIIIGALAAW---CYNKFYNI 156  
QY 255 RLMTGADFL---ITPVLTLTLTGFLTAIGPAMRWVGDLLAHGLQ---GLYDFGGPV 306  
Db 157 NLPSYLGFFAGKRFVPIV-MATTSFI--LAPPMALIW--PTIOTGLNATSTGLDSDNTGV 211  
QY 307 GGLLGLVYSPIVITGLHQSF-----PPIELFNOG-----G 339  
Db 212 AVFLFGFIKRLIPFGLHHIFHAPFWFEGSKNAAGEIIHGDDRIETIEQIREGAHLTAG 271  
QY 340 SFIFATASMANIAQGAACLAFFFLAKSEKLGAGAGVSA---VLGITEPAIFGVNLR 395  
Db 272 KMQGEFFVMMFGLPAAALAIYHSKPKENKVVAGLMSAALTSLTGTITPLEFSFLV 331  
QY 396 LWPFFYIGTAAIGGALIALFDIKAVAGLAAAGFLGVVSIDA--PDMVMF-----LVCA 447  
Db 332 APLLFFIHAVLDGLSFLTLVLDVHLGYTFSGGFDYVLLGVLPNKQTQWMLVIPVGLVYA 391  
QY 448 VVTF-----VIAFGAAIAYGLY-----LVR- 467  
Db 392 VIYFVFRFLIVKLKYKTPGREDKQSQAVTASATELPYAVLEAMGGRANIKHLDACITRL 451  
QY 468 -----RNGSIDPDATAA 479  
Db 452 RVEVNDKSKVDVPGLKDLGASGVLEVGNMQAIFGPKSDQIKHEMQQIMNGQVVENPTM 511  
QY 480 PVPACTTKAEAPAEAFESNDSTIIQAPLTGEAIALSSVSDAMPASGKLGSGVAIVPTKGQ 539  
Db 512 EDDKDETIVVAEDKSAISELSHVHAPLTGEVTPLEVPDQVFEKMMGMDGIAIKPSQGE 571



```
Query Match          9.5%; Score 319; DB 9; Length 687;
Best Local Similarity 21.6%; Pred. No. 6.3e-20;
Matches 149; Conservative 110; Mismatches 236; Indels 194; Gaps 28;

QY 103 FSRVAVLVADIVPILVGGGLMAINNVLVADLFGPSQSLVEMFPQISG-----VAE 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 FGQQRIGKALMLP-VAILPAAGILLAFGNAM-----HNEQLVFIAPWLKNDIIVMISS 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 MNLMASAPFALPVL--VGFTATKRFNGN--EFLGAGIGMAMVPTTL-----VN 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 VNEAGQVVDNPLFLFAVG-TALGLAGDGVAAALVGLYLMNATMGKVLHITDIDIF 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 GYDVAA---TMTAGEMPMWSLGLDVAQAGYOGTVLPVLVSVWILATIEKFLHKLMTA 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 SYAKGAKELSOAKPEPAHALVLGIPTLQTVGGIIMGALAAW---CYNKFYNITLPPEL 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 DEL---ITPVLTLTGLTFTIATIGPMRWVGDLLAHGLOGLYDF---GGPVGGLLFG 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 GFAGKRFVIVTSVA-----IATGVLLSFPAWPIQDGLNLSNELLNKNLTITTFIFG 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 LVSPVITGLHQSF--PPIEL--FNOGGSFIFATAS--MANTAQG-----354
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 ITERSLLPGLHIFSPWFEGSYTHAGELVRGQRIWMAQLKDGVPFAGATTGK 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 -----AACLVFFLAKSEKLV---GLAGASVSAYL-GITEPAIF-----PAIFG 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 YPFMFGLPARAFIYKNARPKKVVGGMLSLAGLTAELTGITEPLESFLVAPVLVG 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 VNLRWPPYIGTAAAGALIALFDIKAVAGLGAAGFL---GVVSDADPMVMFLVC 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 IHVLL-----AGTSFL---VHMLLVGKIGMTFSGGFDYILYGLLNWDRSHALLVIPV 399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 447 AVV-TFVIAFGAAIYGLVVRNGSIDPD-----ATAAPVP-----482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 GIVYAIYVFLDFAIRKFKLTPGREDEETEIRNSSAKLPFDVLDAMGGKENIKHLDA 459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 483 -----AG-----486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 460 CITRLRVEVVDKSVDAIGKALGASGVLEVGNMKAIFGPKSDQIKHDMKIMSEITK 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 487 KARAEPAEFSND-----STTQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKG 538
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 PSETTTEEMSDPEVHVEALGTDTIYAPGIGIIPLSEVPDQVFAGKMGDGVGFPEKG 579
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 539 QLVSPVSGKIYVAPPSGHAFVTRKAEAGSNVDILMHIGFDVNLNGTHFNPLKKGDEV 598
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 580 EIVAPDGIYKTIPTPKHAIGL-----ESESQVEVLHIGIDIVKLNGEGFESLINVDEKV 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 599 KAGELCEFDIDAIAKAGYEVTTPVIVSN 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 TQSQPLMKVNLAYLKAHAPSIVTPMIITN 664
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-741-669-330
; Sequence 330, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 330
; LENGTH: 648

; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-330

Query Match          9.28; Score 308.5; DB 9; Length 648;
Best Local Similarity 22.1%; Pred. No. 5.2e-19;
Matches 149; Conservative 85; Mismatches 241; Indels 199; Gaps 21;

QY 152 ISGVAEMINLMASAPFALPVLVGTATKRFGGNEFLG-AGIGMA--MVFPFVLVNGYDVA 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ILGFFQRLGRALQLPIAFLPVA---ALLRFGQPDLLNVAFIAQAGGAIFDNL-----A 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 ATMTAGEMPMWSLGLDVAQAGYOGTVLPVLVSVWILATIEKFLHKLMTADFLITPVL 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 LIFAIGVASSWS---KDSAGAAALAGAVGVFLTKAMVTINPEIN---MGVLAGIITGLV 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 -----TLLTGTFTF-----IAIGPMRWVGDLLAHGLOGLYDF-- 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 GGAAYNRWSDIKLPDFLSFGGKRFVPVPIATGFFCLVLAALFGYVWPPVPHAIHAGGWI 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 -GGPVGGLGLVSPVITGLHQSFPII-----EELFNQG-----G 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 SAGALSGSIFGINRLIPTGLHQLVNTIAWFQIGFTNAGTVFHDGIDNRFYAGDGTAG 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 340 SFIPATASMANIAOGAACLAVFFLAKSEK---LKLAGASVSAYL-GITEPAIF----- 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 MFMSGFPIMMFGILPGAALAMYFAKERRPMVGMGLLSVAVTAFLTGVTPELFLFMEFL 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 391 -----GVNLRWPPYIGITAAIGGAL-----413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 APLLYLLHALLTGISLFAVATLLGIHAGFSFSAIDAVALMYNLPAASONMYMLLVMGVIF 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 414 -----IALFDIKA-----VALGAA-----427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 FAIVFVFSLVIRFENLKTGPREKDEIVTEEANSNTEBGLTQLATNYTAAYVGGTDNLK 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 -----GFLGVSDADPMVMFLVCAVTVFVIAFGAIA 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 AIDACITRLRUTVADSARVNDTMCKRLGASGVVKNKQTIQIVVGAKEISGDAMKKVVA 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 461 YGLVLRNGSIDPDATAAPVPACTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDA 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 RG-PYAAASABATP-ATAAPVAKPOAVPNVAVSIAE-----LVSPITGDVVALDQVPDE 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 MFASGKLGSAIVPTKGQLVSPVSGKIYVAPPSGHAFVTRKAEAGSNVDILMHIGFD 580
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 AFASKAVGDGVAVKPTDKIVVSPAAGTIVKIFNTINHAFLCLETE-----KGAIEIVVHMGIDT 574
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 581 VNLNGTHFNPLKKGDEVKAGELCEFDIDAIAKAGYEVTTPVIVSNYKKTGPNVTYGLG 640
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 575 VALEKGFRLVEEGAQVSAQOPILEMDLDLYLNANARSMISPVVCSNIDDFSGLIIRKAQG 634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 641 EIEAGANLLNVAKK 654
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 635 HIVAGQTPLYEIKK 648
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-09-815-242-10089
; Sequence 10089, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; FILE REFERENCE: ELITRA.011A
```

Query Match	9.2%	Score	308.5	DB	9	Length	648
Best Local Similarity	22.1%	Pred. No.	5.2e-19				
Matches	149	Conservative	85	Mismatches	241	Indels	199
Gaps	21						
Qy	152	ISGVAEMTNLMASAPFAPLPVLVGFTTKRFGGNEFLG-AGICMA--MVFPPLVNGYDVA	208				
Db	3	ILGFQRGRALQLPIAIVLPA--ALLRFGQDLLNNAFAIAQAGGAIFDNL-----A	53				
Qy	209	AWTAGEWMPMSLFGDLVQAQAGYOGTVLPVLVVSWSLTIETKFLKRLMGTADFLITPVL	268				
Db	54	LIFAICGVASSMS--KDSAGAAALAGAVGVFLTKAMVINPEIN--MGVLAGIITGLV	107				
Qy	269	-----TLTLTGFLTF-----IAIGPAMRWVGDLLAHGQGLGYDF--	302				
Db	108	GGAAYNRWSDIKLPDLFSFGKRFVEPIATGFCFLVLAATFGVYVPPVQHAHAGGENTV	167				
Qy	303	-GGPVGGLLFGVLVSPIVITGLHGSFPPI-----EULEFNQG-----G	339				
Db	168	SAGAGSGTGFSGINLLPTGLHVLNTIAWFOIGFTNAAGTVFHGDIRRVYAGDGTAG	227				
Qy	340	SFEIATASMANIAQGAACLAVFVFLAKSEK--LKGLAGASGVSAVL-GITEPAIF-	390				
Db	228	MFMSGFFPIIMFGLPGAALAMYFAKPRRPWVGMLLSVAVTAELTGTEPLELFWEL	287				
Qy	391	-----GVNLRURWPFYIGITAAIGAL-----	413				
Db	288	APLYLLHALLTGISLFVATLLGHAGFSFSAGAIDYALMYNLPAASQWVMLLYMGVIF	347				
Qy	414	-----IALDPDKA-----	427				
Db	348	FAIFYVVSFLVRMENLTPGREDKEDEIVTEEANSNTEEGLIQLATFNIAAVGDTDNLK	407				
Qy	428	-----GFLGVWSIDAPDMVMFLVCVVTFVIAFGAATA	460				
Db	408	AIDACITRLRLVADSARVNDTMCRLKAGSVQVXNKRQTIQTVGAKASIGIDAMKKVYA	467				
Qy	461	YGLYVRRNGSIDDPATAAPYAGTTRKAEPAPAEFSNDSTIIQAPLTGEATALLSSVSDA	520				
Db	468	RG-PVAAASAETP-ATAAPVAKPQAPVNAVSTAE-----LVSPITGDVVALDOVPDE	518				
Qy	521	MFASKGLGSGVAIVPTKQLVSPVSGKIWAFFPSGHAFVAVRTKAEDGSNNVDILMHIGFDT	580				
Db	519	AFASKAVGGVAVKPTDKIVVSPAAGTIVKIFNTNHAFCLETE---KGAEIVVHMGIDT	574				
Qy	581	VNLNGTHFNPLAKQGDYKAGELLCEFDIDAIKAAGYEVTPPIVVSNNYKKTGPVNTYIGLG	640				
Db	575	VALEGHGFRLVEEGAQVSAGQPIEMDLDYLNANAKSMISPVVVCSSNIDDFSGLIIKAOG	634				
Qy	641	EIEAGANLLNVAKK	654				